

Universidade de Lisboa
Faculdade de Ciências
Departamento de Biologia Vegetal



“Identificação, caracterização e mapeamento físico de retroelementos em *Quercus suber* L.”

Em associação com
Universidade Técnica de Lisboa
Instituto Superior de Agronomia



Margarida Morais de Carvalho

**Mestrado em Biologia Celular e Biotecnologia
2007**

Universidade de Lisboa
Faculdade de Ciências
Departamento de Biologia Vegetal



**“Identificação, caracterização e
mapeamento físico de retroelementos em
Quercus suber L.”**

Em associação com
Universidade Técnica de Lisboa
Instituto Superior de Agronomia



Margarida Morais de Carvalho

Dissertação orientada pelo Prof. Doutor Rui Gomes e pela Dra.
Margarida Rocheta
Mestrado em Biologia Celular e Biotecnologia
2007

Índice

	Pág.
Agradecimentos	i
Lista de abreviaturas	ii
Resumo	iii
Abstract	iv
 Capítulo 1. Introdução	
1. Elementos genéticos móveis	1
1.1 Retrotransposões e os retrovírus	3
1.1.2 Qual a origem dos retrotransposões?	5
1.2. Localização genómica de retrotransposões	6
2. Objectivos	8
 Capítulo 2. Materiais e métodos	
1. Material vegetal	10
2. Análise de DNA	10
2.1 Extracção de DNA genómico	10
2.2 Visualização de DNA por electroforese em gel de agarose e sua quantificação	10
2.3 Isolamento de LTR's 3' de retrotransposões <i>Ty1-copia</i> em <i>Q. suber</i>	10
2.3.1 Digestão de DNA genómico com enzimas de restrição	10
2.3.2 Purificação de DNA	11
2.3.3 Preparação do “template DNA” - ligação de adaptadores	11
2.3.4 Amplificação de DNA por PCR	11
2.4 Isolamento de <i>RVT's</i> de retrotransposões <i>Ty1-copia</i> em <i>Q. suber</i>	12
2.4.1 Amplificação de DNA por PCR	12
2.4.2 Isolamento de fragmento(s) de DNA a partir de um gel de agarose	12
2.5 Clonagem e sequenciação	13

2.5.1	Preparação de células competentes (<i>E. coli</i> estirpes DH5a e XL1B)	13
2.5.2	Preparação do “insert”- ligação de produtos de PCR	13
2.5.3	Transformação	13
2.5.4	Seleção das colónias transformadas por PCR	14
2.5.5	Extracção de DNA plasmídico e sequenciação	14
3.	Análise bioinformática	14
4.	Análise Citogenética Molecular	15
4.1	Silanização de lâminas	15
4.2	Pré-tratamento do material biológico	15
4.3	Técnica de drop em raízes de <i>Quercus suber</i>	15
5.	Análise de DNA pelo método de C-PRINS (Cycling primed <i>in situ</i> DNA labelling)	16
5.1	Pré-tratamento do material	16
5.2	Preparação e adição da mistura de PCR	16
5.3	C – PRINS (Cycling Primed <i>in situ</i> DNA Labeling)	17
5.4	Lavagens pós-reacção	17
5.5	Deteção dos locais onde ocorreu amplificação	17
5.6	Visualização, captação e processamento de imagens	18
6.	Hibridação <i>in situ</i> fluorescente (FISH)	18
6.1	Preparação da sonda	18
6.2	Pré-tratamento do material	18
6.3	Preparação da mistura de hibridação	19
6.4	Desnaturação e hibridação	19
6.5	Lavagens pós-hibridação	19
6.6	Deteção dos locais de hibridação	20
6.7	Visualização, captação e processamento de imagens	20

Capítulo 3. Resultados

1.	Isolamento de sequências de <i>RVT</i> 's de retrotransposões <i>Ty1-copia</i> em DNA genómico de <i>Quercus suber</i>	21
----	--	----

2. Padrões de distribuição de <i>RVT</i> 's de retrotransposições <i>gypsy</i> e <i>copia</i> -like no genoma de <i>Quercus suber</i> por C-PRINS (PCR <i>in situ</i>)	24
3. Isolamento de sequências “ <i>env</i> -like” em retrotransposições <i>Ty1-copia</i> em DNA genómico de <i>Quercus suber</i>	26
4. Padrão de distribuição de sequências “ <i>env</i> -like” no genoma de <i>Quercus suber</i> por PCR <i>in situ</i>	29

Capítulo 4. Discussão

1. Retrotransposições LTR em plantas: transmissão vertical, horizontal ou ambas?	30
2. Contribuição das sequências de <i>RVT</i> 's e “ <i>env</i> -like” para a organização do genoma de <i>Q. suber</i>	35
3. Conclusões finais	37
Referências bibliográficas	38

Agradecimentos

Neste espaço particular cabe-me agradecer a todos aqueles que, de uma forma ou de outra tornaram possível o meu curso de Mestrado.

Em primeiro lugar quero agradecer à doutora Margarida Rocheta, por me ter mostrado um novo mundo e por me ter permitido fazer parte dele. Pela paciência nos meus momentos de incompreensão, de tensão e mesmo de descrença...afinal, sempre consegui dar a volta por cima e ultrapassar os inúmeros obstáculos que se nos depararam. Assim, tudo isto tem um valor muito diferente...

Em seguida quero agradecer ao professor Rui Gomes por ter aceite ser meu orientador e pela sua pronta disponibilidade.

A todas as pessoas que me acolheram no ISA, quer no laboratório de genética, quer fora dele. Foi também por vossa causa que este ano e meio de trabalho foi tão gratificante para mim e que, dessa forma, se tornou ainda mais leve. À professora Wanda, pela recepção calorosa e por ter permitido que tudo isto fosse possível ao receber-me no seu laboratório. À professora Leonor, pela incansável paciência e encorajamento nos momentos mais difíceis. À “minha” Teresinha... pela pronta disponibilidade em ensinar-me muito do que hoje sei de citogenética, mas também por aturar a minha energia na bancada durante muitas e muitas horas. É obra Teresinha! Miguel... que te hei-de dizer? Colega de bancada...por si só já diz muito, mas foste mais do que isso. Ainda vou ter saudades das nossas lutas pelo espaço, pelos bancos, pelas canetas...ou não, porque afinal já as tenho!Ana Luisa...a “minha” companheira desportiva e sem dúvida um apoio em alturas menos boas, além de cúmplice nos momentos de risota. Márcia, pelo acolhimento caloroso e pelos momentos de diversão, dentro e fora do ISA. Aos “miúdos”, João e Larginho por muitos dos melhores momentos que passei no ISA. Augusta, Guida, Manuela...sem vocês o que seria das barrigas da miudagem? Há que manter o pessoal em forma para ir buscar terra e vasos...e pregar partidas! Ao João, pela ajuda nos dias mais trabalhosos em que a luz do dia há muito se tinha posto...pela amizade, companheirismo e por ter aberto muitas janelas ainda fechadas na minha cabeça. A todos eles, e outras pessoas que não estão aqui descritas, mas que fizeram parte do meu dia-a-dia neste instituto, o meu muito obrigada.

Aos meus amigos, pela paciência e pela força, acabei!!!Mas também com o vosso apoio...ao Zé, que em situações de stress me ajudou SEMPRE!!Incansável.Obrigada.

E finalmente, à minha família. Sem vocês tinha sido indubitavelmente muito mais difícil, mas o vosso apoio esteve e está sempre presente na minha vida. Obrigada por tudo...vocês sabem o quê.

Lista de abreviaturas

bp- pares de bases

BSA - Albumina de soro de bovino

DAPI - 4',6-diamidino-2-fenilindol

dUTP – Desoxitimidinatrifosfato

dNTP – Desoxitrifosfatos de nucleósidos

cDNA- DNA complementar

DNA - Ácido desoxiribonucleico

env - envelope

EDTA - Ácido etilenodiaminotetracético

FITC- isotiocianato de fluoresceína

g, mg, µg - grama, miligrama, micrograma

h, min - Hora, minutos

HEPES - Ácido 4-(2-hidroxietil)-1-piperazinoetanosulfónico

INT- integrase (endonuclease)

L, ml, µl - litro, mililitro, microlitro

LB - Meio Luria-Bertani

M, mM, µM- Molar, milimolar, micromolar

NOR - Região organizadora de nucléolo

°C - Graus Celsius

PBS - Solução tampão de fosfato; primer binding site

PPT- popypurine tract (zona rica em purinas)

PR – protease

rDNA - DNA ribossomal

RNA - Ácido ribonucleico

RNAse – Ribonuclease

RnaseH – ribonuclease H

Rpm – rotações por minuto

rRNA - RNA ribossomal

RVT- Polimerase de DNA dependente de RNA; transcritase reversa

SDS - Dodecil sulfato de sódio

SSC - Citrato de sódio salino

TAE – solução tampão de tris-acetato e EDTA

U – unidade enzimática

Resumo

Os retrotransposições LTR são elementos genéticos móveis ubíquos nos organismos eucariotas, em particular no reino vegetal, o que traduz a sua ancestralidade. À semelhança dos retrovirus, transpõem através de um intermediário de RNA, que é convertido em cDNA pela transcriptase reversa (RVT) e RnaseH, sendo posteriormente integrado no genoma hospedeiro. Consoante a sua organização genética e homologia entre as suas sequências classificam-se em *Ty1-copia* ou *Ty3-gypsy*, à semelhança dos elementos tipo de *Drosophila melanogaster*. Recentemente foi descrita a presença do gene *env*, característico de retrovirus e envolvido na sua capacidade infecciosa, em retrotransposições LTR *Ty1-copia* e *Ty3-gypsy*, o que levantou a hipótese de alguns destes elementos serem retrovírus endógenos. Adicionalmente a elevada homologia entre retrotransposições de espécies evolutivamente distantes sugere que tenha ocorrido transmissão horizontal durante a evolução dos mesmos. Em plantas estes elementos são constituintes maioritários dos genomas, alcançando 80% do genoma nuclear de milho. A espécie *Quercus suber* é uma das mais importantes espécies de árvores em Portugal, a nível económico e ecológico, mas poucos estudos têm sido feitos acerca dos retroelementos e da sua importância na estrutura/evolução do genoma.

Neste trabalho inicia-se o estudo dos retrotransposições LTR em sobreiro, pela identificação e isolamento de sequências de transcriptases reversas (RVT's) e "env-like" em retrotransposições *Ty1-copia*. Adicionalmente estuda-se a distribuição física de RVT's *copia* e *gypsy-like* e de uma sequência "env-like" no genoma desta espécie.

Os resultados obtidos revelam a presença de retrotransposições *copia* e *gypsy-like* em sobreiro, além de sequências "env-like", dispersos pelo genoma e ausentes de regiões heterocromáticas à excepção do elemento *gypsy-like*, que exhibe uma marcação mais intensa no *locus* de rDNA 45S NOR1 e em centrómeros. Análises filogenéticas destas sequências permitiram concluir que embora a transmissão vertical seja preponderante na dispersão destes elementos, parece ter ocorrido transmissão horizontal em alguma altura da evolução dos mesmos.

Palavras-chave: *Quercus suber*, retrotransposições LTR, filogenia, transcriptase reversa, *envelope*

Abstract

LTR retrotransposons are ubiquitous mobile genetic elements in eukaryotic organisms, specially in plant kingdom, which represents their ancestrality. As retroelements, like retrovirus, they transpose through an RNA intermediate that is converted to cDNA by reverse transcriptase (RVT) and RNaseH, being then suitable for integration on the host genome. According to their genetic organization and sequence similarity they are classified as *Ty1-copia* or *Ty3-gypsy*, much like the type elements of *Drosophila melanogaster*. Recently it has been reported the presence of the gene *env*, characteristic of retrovirus and involved in their infectious capacity, in LTR retrotransposons *Ty1-copia* or *Ty3-gypsy*, which raised the hypothesis of some of these elements being in fact endogenous retroviruses. Additionally the strong homology between retrotransposons of distant evolutionary species suggests that horizontal transmission has occurred during the evolution of these elements. In plants retrotransposons are major constituents of genomes, comprising 80% of the nuclear genome of maize. The *Quercus suber* species is one of the most important tree species in Portugal, both economically and ecologically. Nevertheless only but a few studies have been made about retrotransposons and its importance in the structure/evolution of the genome.

In this work is strated the study of LTR retrotransposon in cork oak, through the identification and isolation of transcriptase reverse and “*env*-like” sequences in *Ty1-copia* retrotransposons. It is also studied the physical distribution of *copia* and *gypsy*-like RVT's and an “*env*-like” sequence in the genome of this species.

The achieved results show the presence of *copia* and *gypsy*-like retrotransposons in cork oak, in addition to *env*-like sequences, spreaded through the genome and absent from heterocromatic regions, except for the *gypsy*-like element which exhibits a stronger signal in the rDNA 45S *locus*, NOR1 and in centromeres. Phylogenetic analyses provided evidence that although vertical transmission is the main force driving the spread of these elements, it seems that horizontal transmission of retrotransposons has occurred at some point of their evolution.

Keywords: *Quercus suber*, LTR retrotransposons, phylogeny, reverse transcriptase, *envelope*

Capítulo 1. Introdução

1. Elementos genéticos móveis

Os elementos genéticos móveis são segmentos cromossômicos com capacidade de mobilização no genoma e foram descobertos numa planta superior, o milho (*Zea mays*) por Barbara Mcclintock, em 1949. São elementos ubíquos, estando presentes numa grande variedade de organismos, incluindo bactérias, fungos, plantas e animais, o que traduz a sua ancestralidade. Consoante o mecanismo pelo qual se movem no genoma, podemos dividir estes elementos em dois grupos: os transposões e os retroelementos. Os transposões movem-se no genoma hospedeiro à custa de um mecanismo de excisão e posterior integração noutra local do genoma pela acção da enzima transposase, por eles codificada. Os retroelementos movem-se através de um intermediário de RNA que é sujeito a transcrição reversa pela enzima transcritase reversa e o cDNA correspondente pode posteriormente ser integrado no genoma hospedeiro. Deste último grupo fazem parte os retrotransposões, os retrovírus de vertebrados, os elementos DIRS1 e BEL e os caulimovírus (Malik et al. 2000). Os retrotransposões constituem a maior e mais dispersa classe de elementos transponíveis eucariotas, sendo agrupados em duas grandes famílias consoante a presença ou ausência de repetições longas terminais (LTR's), retrotransposões LTR e não-LTR (LINE's (long interspersed repetitive elements) e SINE's (short interspersed repetitive elements) respectivamente. As LTR's destes elementos podem variar de 100bp até mais de 5kb e não codificam para proteínas contendo, porém, promotores e terminadores associados à transcrição dos retrotransposões. Após a LTR 5' localiza-se o domínio PBS (primer binding site) e antes do início da LTR 3' encontra-se a região PPT (polypurine tract), elementos *cis* envolvidos na transcrição destes elementos (Figura 1.1). No interior destes domínios localiza-se a zona codificante, especificada por genes semelhantes aos dos retrovírus, nomeadamente os genes *gag* e *pol*. O gene *gag* codifica para proteínas estruturais envolvidas na maturação do RNA do retrotransposão e na montagem de partículas semelhantes a vírus (VLP- vírus like particules), enquanto que o gene *pol* codifica para proteínas com função catalítica: uma protease (*PR*), uma transcritase reversa (*RVT*), uma *RNase H* (*RH*) e uma integrase (*INT*) (Varmus and Brown 1989). Com base na ordem dos domínios deste gene os retrotransposões LTR dividem-se em *Ty1-copia* e *Ty3-gypsy*, à semelhança dos elementos tipo de *Drosophila melanogaster*. A diferença entre eles reside na posição ocupada pela INT relativamente à RVT, sendo PR-INT-RVT-RNaseH nos elementos *copia* e PR-RVT-RNaseH-INT nos *gypsy* (Figura 1.1).

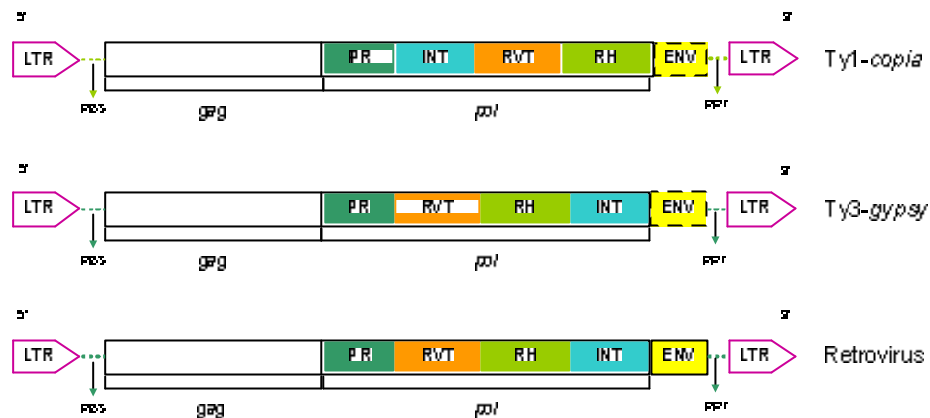


Figura 1.1- Esquema representativo da estrutura básica de um retrotransposão LTR, *Ty1-copia* ou *Ty3-gypsy*, e de um retrovírus. Todos são constituídos por uma região codificante especificada pelos genes *gag* e *pol*, flanqueada por elementos *cis* envolvidos na sua transcrição, PBS e PPT que são limitados exteriormente por duas LTR's. A diferença entre retrotransposões LTR e retrovírus consiste na presença de um gene, *env*, característico dos retrovírus, que poucos retrotransposões contêm. No entanto, quando presentes localizam-se a montante da LTR 3' e a jusante do gene *pol*, a mesma localização que se encontra nos retrovírus. Dentro dos retrotransposões LTR, a ordem dos domínios RVT e INT do gene *pol* pode ser utilizada como ferramenta taxonómica, sendo RVT-RH-INT nos elementos *Ty3-gypsy* e INT-RH-RVT nos *Ty1-copia*.

Por outro lado, a análise filogenética de sequências de RVT's, a enzima com maior grau de conservação nestes elementos, em ambas as famílias de retrotransposões confirma que os elementos *Ty1-copia* e *Ty3-gypsy* formam duas linhagens evolutivamente distintas. Actualmente pensa-se que os mecanismos de transcrição, replicação e integração sejam os mesmos para os retrotransposões LTR e retrovírus. A transcrição caracteriza-se pela síntese de um único mRNA que funciona como molde para a síntese de novas cópias do retrotransposão e para a tradução das suas proteínas. O mRNA, análogo ao RNA genómico retroviral (Boeke 1989), é traduzido numa poliproteína que é clivada em péptidos funcionais pela protease codificada no próprio retrotransposão. A RVT e a RNaseH são responsáveis pela síntese do cDNA correspondente e degradação da molécula de mRNA, respectivamente. A INT é necessária à integração do cDNA no genoma hospedeiro. A nova molécula do retrotransposão é uma molécula circular de dupla cadeia e com LTR's idênticas. Para ser introduzida no genoma, a molécula do retroelemento é linearizada, através de um corte entre as suas LTR's, pela enzima integrase que efectua também um corte na cadeia de DNA do genoma hospedeiro, local onde é inserida a nova cópia do retroelemento. As LTR's têm sequências idênticas aquando da sua síntese, consequência do mecanismo de replicação, e o acumular de mutações ao longo dos ciclos celulares permite a sua utilização na datação da introdução do retroelemento no genoma. Como resultado deste mecanismo de transposição de elevada capacidade invasiva, os

retrotransposições podem originar um elevado número de cópias e aumentar de forma significativa o tamanho do genoma hospedeiro, tornando-se um dos seus constituintes maioritários (Bennetzen 2002; SanMiguel et al. 1996a). Ao se inserirem em zonas próximas de genes (Kunze et al. 1997), os retrotransposições podem afectar a expressão genética, como é o caso do alelo B-Bolivia do gene *b1* em milho, que possui na sua proximidade uma sequência de um retrotransposição (Selinger and Chandler 2001). Noutros casos os retrotransposições inserem-se dentro de genes causando mutações que na sua maioria têm efeitos deletérios. Como tal não é de surpreender a existência de mecanismos de regulação da sua actividade transposicional, quer por parte do hospedeiro, quer por factores codificados pelos próprios retrotransposições, provavelmente para minimizar os efeitos negativos, garantindo a sobrevivência do hospedeiro e, conseqüentemente, a sua. Um dos passos de maior importância no controlo da mobilidade dos retrotransposições é a transcrição. Em retrotransposições LTR o controlo transcricional envolve sequências reguladoras em *cis* que estão normalmente localizadas na LTR 5' ou a jusante da zona codificante, em locais não traduzidos (Kumar and Bennetzen 1999). Muitos dos retrotransposições estudados até à data são activados transcricionalmente por factores de stress biótico e abiótico, envolvendo factores de transcrição que se ligam à LTR 5', a pequenas sequências palindrómicas (Kumar and Bennetzen 1999).

1.1 Retrotransposições e os retrovírus

Os retrotransposições LTR e os retrovírus são retroelementos e como tal partilham algumas semelhanças. No entanto os retrovírus possuem uma ORF adicional, localizada a jusante do gene *pol* e a montante da LTR 3' , que codifica para o gene *env* (Figura 1.1). Este gene codifica para glicoproteínas transmembranares envolvidas no reconhecimento celular dos viriões, facilitando a fusão entre o invólucro lipídico do virião e a membrana da célula hospedeira, permitindo a entrada do genoma retroviral na célula (Temin 1980). Até recentemente pensava-se que os retrovírus estavam restringidos aos vertebrados mas em 1994 foi detectada a presença de um retrovírus num invertebrado, *Drosophila melanogaster* (Kumar and Bennetzen 1999; Song et al. 1994).

Os retrotransposições LTR, como referido no ponto 1, dividem-se em duas famílias consoante a ordem dos domínios codificados no gene *pol* (Figura 1.1) sendo a ordem dos elementos *Ty3-gypsy* a mesma que se encontra nos retrovírus, isto é, a RVT precede a INT. Análises filogenéticas das sequências de RVT's destes retroelementos mostram que a homologia os elementos *Ty3-gypsy* estão mais próximos dos retrovirus, relativamente aos elementos *Ty1-copia* o que levou a uma reclassificação dos retrotransposições LTR: *Ty3-*

gypsy como *Metaviridae* e *Ty1-copia* como *Pseudoviridae* (Hull 2001). Nos últimos anos têm sido identificados retrotransposições LTR em plantas que contêm uma ORF adicional, localizada a jusante do gene *pol* e que possui uma elevada homologia com o gene *env* dos retrovírus. Dos retroelementos identificados a grande maioria são do tipo *Ty3-gypsy* dos quais são exemplo a família *Athila* de *Arabidopsis thaliana* (Wright and Voytas 1998a) com os homólogos *Cyclops* (Chavanne et al. 1998) e *Calypso* (Peterson-Burch et al. 2000) em ervilha e soja, respectivamente. Mais recentemente foi descoberto um novo elemento em ervilha designado *PIGY* (Neumann et al. 2005). Em número bastante mais reduzido foram identificados alguns retrotransposições *Ty1-copia* : *SIRE-1* de soja (Laten et al. 2003; Laten et al. 1998) e a família de homólogos em *Arabidopsis thaliana* *Endovir* (Peterson-Burch et al. 2000). No entanto a função do gene *env* em plantas permanece desconhecida. A parede celular constitui um obstáculo à propagação de partículas pelo mecanismo de fusão membranar, o que poderá explicar os poucos casos descritos de infecção retroviral nas plantas. Ainda assim, existem alguns retrovírus em plantas, em particular alguns elementos das famílias *Bunyaviridae* e *Rhabdoviridae* (van-Regenmortel et al. 2000; Vandenheuvel et al. 2002). Estes vírus formam o seu invólucro lipídico a partir do sistema endomembranar da célula hospedeira e ficam retidos dentro desta até que um organismo invertebrado, por exemplo, os transporte para outras plantas durante o processo de ingestão. Outra possibilidade de mobilização é através dos plasmodesmos existentes nas células vegetais. Estes são constituídos por proteínas (conexinas) e perfis de retículo endoplasmático que ficam retidos aquando da formação da parede celular e que permitem a ligação do citoplasma de células contíguas. Um dos vírus que utiliza esta via para a sua propagação na planta é o vírus do mosaico do tabaco. Este codifica para um péptido de sinalização que é responsável pelo aumento do diâmetro do poro do plasmodesmo, controlando o limite de exclusão de tamanho molecular (Carrington et al. 1996). Foram já descritos alguns retrotransposições LTR *Ty3-gypsy* cujo gene *env* é transcrito e cuja tradução conceptual dá origem a uma proteína transmembranar (Neumann et al. 2005). Nestes casos torna-se difícil distinguir entre um elemento *Ty3-gypsy* e um retrovírus, uma vez que a ordem nos seus domínios é a mesma (Bennetzen 1996; Kumar 1998). Exemplo disso é o retrotransposição *Gypsy* de *Drosophila melanogaster*, cujo gene *env* é funcional e, consequentemente levou à sua reclassificação deste elemento como retrovírus (Song et al. 1994). Em plantas têm sido descritos casos semelhantes, por exemplo, pensa-se que os elementos *Athila4* e *Calypso* de *Arabidopsis* e soja, respectivamente, definem uma linhagem de retrovírus endógenos (Wright and Voytas 2002). No retroelemento *SIRE-1* foi detectada a presença de uma terceira ORF, cuja tradução conceptual produz um péptido de 22kD (ou 28kD por “readthrough” do codão stop), que se localiza na zona 5’ do gene *env* e na orientação antisense. Embora não haja muita informação acerca desta sequência, a estratégia de

expressão em ambos os sentidos, ambissense, é comum em vários vírus de plantas, incluindo o *Toposvirus*, pertencente à família *Bunyaviridae*, sobretudo para a expressão de proteínas envolvidas no movimento do vírus (Mumford et al. 1996). Apesar disso, quer nos retrovírus quer nos retrotransposições LTR nos quais foi identificado (Song et al. 1994) o gene *env* revela um baixo grau de conservação da sua sequência (Laten et al. 1998; Wright and Voytas 1998b).

1.1.2 Qual a origem dos retrotransposições?

Várias hipóteses têm sido sugeridas para a origem dos retrotransposições LTR e dos retrovírus. A ubiquidade dos retrotransposições LTR em organismos eucariotas, em particular nas plantas, é um facto para o qual foram propostas algumas teorias. Uma delas afirma que a ubiquidade destes elementos resulta da sua ancestralidade e que a sua existência data de uma altura prévia à divergência das espécies que as contêm. Análises filogenéticas baseadas em sequências de genes de RVT e INT quer de elementos *Ty1-copia* quer *Ty3-gypsy* em plantas forneceram provas consistentes de transferência vertical de retrotransposições, com algumas excepções (Flavell et al. 1992; Matsuoka and Tsunewaki 1999; Vernhettes et al. 1998). De acordo com esta hipótese os retrotransposições partilham uma maior homologia com elementos isolados na mesma espécie ou em espécies com origem evolutiva próxima, relativamente a elementos de espécies distantes. No entanto, existem casos descritos em que se observa uma maior homologia entre famílias de elementos isoladas em espécies taxonomicamente distantes, o que contraria a teoria referida acima (Wright and Voytas 2002). Como tal pensa-se que a transmissão horizontal possa ter ocorrido numa determinada altura no decorrer da evolução destes elementos. Contudo, estas teorias embora contrastantes não se excluem mutuamente e a evolução e dispersão destes elementos pode resultar de uma conjugação das duas. A proximidade filogenética dos retrotransposições LTR com os retrovírus tem vindo a fornecer dados que poderão esclarecer os processos envolvidos na origem e dispersão dos retrotransposições. Também neste caso são propostas duas teorias. A primeira sugere que os retrotransposições LTR constituem o legado de infecções de retrovírus animais ancestrais (Laten 1999) e que perderam os seus genes de transmissão extracelular por acção de uma pressão selectiva negativa. Esta teoria apoia-se na existência de retrotransposições LTR, *Ty3-gypsy* ou *Ty1-copia*, que foram classificados ou reclassificados como retrovírus endógenos, nomeadamente o elemento *Gypsy* de *D. melanogaster* (Song et al. 1994) e os elementos *Athila4* de *Arabidopsis*, *Calypso* e a família *Diaspora* de soja (Wright and Voytas 2002; Yano et al. 2005). De facto, a expressão do gene *env* envolve o “splicing” do transcrito e os genes *gag* e *pol* são removidos como se se tratassem de intrões, estratégia comum aos retrovírus.

Esta estratégia foi já detectada na retrotransposição LTR *Bagy-2*, de cevada, cujo gene *env* é sujeito a splicing (Vicient et al. 2001). Por outro lado, o mecanismo de transposição dos retrotransposões envolve a síntese de uma molécula de DNA pela RVT a partir de transcritos. Como tal, não é improvável a hipótese de que alguns transcritos aos quais falta o gene *env* tenham sido utilizados como molde para a síntese de novos elementos, que posteriormente foram integrados no genoma, aumentando o número de retroelementos não infecciosos. A maioria das sequências isoladas em gramíneas aparentam não ser funcionais devido à presença de codões stop, alterações nas grelhas de leitura (frameshift) e ainda deleções (Bennetzen 2000). Por outro lado o grau de conservação do gene *env* (Laten et al. 1998; Wright and Voytas 1998b) encontrado em retrotransposões é reduzido. Não desempenhando uma função aparente nas plantas, este gene poderá ter acumulado mutações ao longo da evolução que o transformaram no pseudogene que hoje se observa em alguns retrotransposões. Adicionalmente, muitos retrotransposões são activados face a condições de stress (Grandbastien 1998). Esta observação pode ser explicada ao adoptarmos a perspectiva de um vírus. Uma situação de stress associada a danos mecânicos causados por animais que se estão a alimentar de uma planta sugere que um vector está disponível para ser transferido para o possível hospedeiro, a planta. No entanto, apesar da relação filogenética mais publicada reflectir a teoria de que os retrotransposões LTR têm origem nos retrovírus de animais, a relação inversa não é inconsistente com os dados e argumentos que foram feitos. Esta teoria sugere que os retrovírus derivam de retrotransposões pela aquisição independente do gene *env* (Doolittle and Feng 1992; Pearce 2007; Smith et al. 1992) em vários passos durante a evolução, ou terá aparecido num retrotransposão e terá sido transferido a outros através de recombinação (Kumar 1998; Pearce 2007).

No reino animal os retrovírus são responsáveis por algumas das mais mortais doenças existentes, mas esta patogenicidade não é reconhecida nas plantas. Pode ser importante determinar o que é que na biologia vegetal restringiu ou limitou a patogenicidade dos retrovírus na ausência de um sistema imunitário celular ou humoral identificado.

1.2. Localização genómica de retrotransposões

Os retrotransposões são ubíquos no reino vegetal, detectando-se a sua presença em algas unicelulares, briófitos, gimnospérmicas e angiospérmicas (Flavell et al. 1992). O número de cópias existente por genoma é variável, sendo frequente a existência de um elevado número de cópias em plantas com genomas de grande dimensão. Exemplos clássicos envolvem o caso do milho, em que 50-80% do genoma nuclear é atribuído a estes

elementos (SanMiguel and Bennetzen 1998), mais de 50% do genoma de cevada (Vicient et al. 1999a) e 22% do genoma de arroz (Ma et al. 2004). Em cevada, uma única família de retrotransposões, a família *BARE-1* representa 5% do seu genoma (Vicient et al. 1999a). Um dos principais responsáveis por estes valores é o próprio mecanismo de transposição destes elementos, capaz de originar um elevado número de cópias em pouco tempo. Este mecanismo parece estar na base do paradoxo do valor de DNA C, isto é, a falta de correlação existente entre o tamanho do genoma e a complexidade de um organismo. SanMiguel e Bennetzen (1998) verificaram que a diferença abrupta nos valores de C dentro da família das gramíneas era devida a uma amplificação diferencial dos retrotransposões. Por exemplo, o arroz e a cevada têm aproximadamente o mesmo número de genes mas a diferença entre o tamanho dos seus genomas é atribuída, em parte, à amplificação destes elementos (Vicient et al., 1999). Nesta perspectiva a diferença que se observa em genomas de grande dimensão, como o de fava e milho (1C de cerca de 14.000Mbp e 2500Mbp respectivamente) pode resultar de uma amplificação bem sucedida de retrotransposões (Bennett and Leitch 2005; Pearce et al. 1996a; SanMiguel et al. 1996b) enquanto que genomas de pequenas dimensões, como o de *Arabidopsis* (1C de 130Mbp), podem resultar de uma ausência de transposição (Wright et al. 1996). Paralelamente, foi demonstrada a preferência que alguns elementos exibem pela amplificação em certas espécies: elementos *copia*-like proliferaram em espécies com genomas mais pequenos, enquanto que elementos *gypsy*-like proliferaram em genomas de maiores dimensões, como em *Gossypium* (Hawkins et al. 2006). Até à data foram feitos inúmeros estudos acerca da localização dos retrotransposões LTR no genoma de plantas com o objectivo de conhecer a sua diversidade e assim poder clarificar a sua função na evolução destes genomas,

Dados resultantes de hibridação *in situ* fluorescente (FISH) em diversas espécies vegetais permitiu concluir que, de uma forma geral, parece haver diferença no padrão de distribuição de elementos *Ty1-copia* e *Ty3-gypsy*.

Elementos *Ty1-copia* parecem não ter preferência por domínios genómicos específicos uma vez que foram detectados tanto em zonas eucromáticas, isto é, zonas ricas nas bases G e C, nas quais se localiza a maioria dos genes e que estão associadas a actividade transcricional (Brandes et al. 1997a; Pearce et al. 1996a; Pearce et al. 1996b; Price et al. 2002) como também em zonas heterocromáticas, ricas em A-T e associadas a inactivação transcricional, nomeadamente zonas teloméricas (*Allium cepa*) (Pearce et al. 1996b) e paracentroméricas (*Arabidopsis thaliana* e *Cicer arietinum*) (Brandes et al. 1997a). Experiências de FISH em *Oryza australiensis* mostraram que a família *RIRE-1* é mais abundante na zona pericentromérica que na zona distal de ambos os braços cromossómicos (Uozo et al., 1997 em (Piegu et al. 2006), uma característica partilhada com outras famílias LTR no genoma de *O. sativa* (Jiang et al. 2002). Noutras gramíneas foi detectada a

presença de retrotransposições LTR aglomerados em “clusters”, num tipo de arranjo designado “nested”, ou seja, elementos inseridos uns dentro dos outros. Estes podem ser da mesma família, como é o caso de *BARE-1*, em cevada (Shirasu et al. 2000) ou de famílias diferentes, em milho (SanMiguel et al. 1998).

Elementos *Ty3-gypsy* são igualmente abundantes em cereais, em regiões eucromáticas mas também heterocromáticas, especialmente em zonas centroméricas, como se verificou em milho, centeio, trigo, arroz e cevada mas também em *Arabidopsis thaliana* (SanMiguel et al. 1996a). Em milho, uma sequência de 240kb localiza-se essencialmente em zonas eucromáticas intergénicas enquanto que em arroz uma família de retrotransposições LTR não-autónomos – *Dasheng* - com um elevado número de cópias, parece exibir uma preferência pela inserção em zonas de microssatélites para além de outros elementos transponíveis, que se localizam em zonas pericentroméricas (Jiang et al. 2002).

No filo *Cnidaria*, assim como em algumas espécies de artrópodes, entre as quais *Drosophila melanogaster*, foi identificada uma família de retrotransposições não-LTR designada R2, específica de DNA ribossomal (rDNA), ou seja de zonas de DNA repetitivo codificante (Kojima and Fujiwara 2005; Kojima et al. 2006). Os genes ribossomais são sequências repetitivas abundantes na maioria dos eucariotas e estão organizados em *tandem* (uns a seguir aos outros), estando presentes em múltiplas cópias em cada *locus*. Cada unidade de repetição de rDNA 45S contém sequências codificantes e não codificantes num padrão organizado e engloba os genes que codificam para as subunidades ribossomais 18, 5.8 e 25S, respectivamente. Estes genes estão separados por duas sequências internas transcritas (ITS1 e 2) e uma região externa não transcrita (ETS) que flanqueia cada repetição. As cópias de rDNA 45S localizam-se num ou mais cromossomas e são visíveis citologicamente numa estrutura particular, a região organizadora do nucléolo – NOR (Zoldos et al. 1999).

2. Objectivos

A espécie *Quercus suber* é de extrema importância em Portugal, quer a nível económico para a produção e exportação de cortiça, quer a nível ecológico pela sua percentagem significativa na flora nacional. Poucos estudos têm sido efectuados sobre a temática dos retroelementos e a sua importância na estrutura/evolução deste genoma. Um primeiro trabalho, em fase de conclusão, reporta um elemento completo pertencente à classe *Ty3-gypsy*, identificado e caracterizado no laboratório de Genética (CBAA). Esta dissertação pretende contribuir para o aumento do conhecimento sobre retroelementos LTR e a sua presença no genoma desta espécie. No início deste trabalho assumiu-se que, tal como em

outros organismos eucariotas, no genoma de *Q. suber* estariam presentes sequências de retroelementos LTR do tipo *Ty1-copia*, especialmente pela sua ubiquidade no reino vegetal (Flavell et al. 1992). Como não havia qualquer informação prévia acerca da presença destes elementos transponíveis em *Q. suber*, assim como de sequências do tipo “*env*”, este trabalho teve como objectivo identificar, clonar e sequenciar retroelementos LTR pertencentes à sub-classe *Ty1-copia* assim como sequências “*env-like*”. Através de uma análise comparativa com elementos descritos, será possível identificar tipos diferentes de retroelementos e inferir acerca da sua transmissão. Paralelamente determinou-se e comparou-se o padrão de distribuição de retroelementos *gypsy* e *copia* no genoma de *Q. suber*, assim como de sequências “*env*”.

Capítulo 2. Materiais e métodos

1. Material vegetal

As plantas utilizadas neste trabalho pertencem à espécie *Quercus suber* e são originárias de Alcácer do Sal. Fizeram-se germinar algumas bolotas em areia húmida e uma vez germinadas colocaram-se em vasos, numa mistura de terra e areia. Os vasos foram colocados em estufa à temperatura ambiente e com um fotoperíodo de 16h de luz, 8h de escuridão. Assim que se obtiveram folhas, recolheram-se algumas e conservaram-se a -80°C até à extracção de DNA genómico.

2. Análise de DNA

2.1 Extracção de DNA genómico

Congelaram-se aproximadamente 1,5g de folhas jovens de *Quercus suber* em azoto líquido e trituraram-se em almofariz previamente arrefecido. O material triturado foi colocado em microtubos de centrífuga e isolou-se o DNA utilizando o kit “Dneasy Plant Mini kit Handbook” (Qiagen), seguindo o protocolo que o acompanha.

2.2 Visualização de DNA por electroforese em gel de agarose e sua quantificação

A técnica seguida foi a descrita por Sambrook (Sambrook et al. 1989). O DNA obtido foi separado em gel horizontal contendo 1% de agarose em tampão TAE (Tris-acetato 0,04M, EDTA 0,01M), utilizando 1xTAE como tampão de electroforese. Antes de colocar os fragmentos no gel adicionou-se a cada amostra 1/10 do volume de Loading Buffer (0,05% azul de bromofenol (p/v), 40% sacarose (p/v), 0,1M EDTA (pH 8.0), 0.5% (p/v) SDS - Sigma G2526) e utilizou-se como marcador de tamanho molecular 1Kb plus DNA ladder (Invitrogen), colocando no gel 4µl do mesmo. Após a migração do DNA, o gel foi corado com brometo de etídio (1µg/ml), observado ao transiluminador ($\lambda = 266\text{nm}$, Sigma T-2201) e fotografado num aparelho Gel Doc 2000 (Bio-Rad). A concentração do DNA genómico extraído foi determinada por espectrofotometria a $\lambda = 280\text{nm}$ num aparelho GeneQuant RNA / DNA Calculator, Pharmacia.

2.3 Isolamento de LTR's 3' de retrotransposições *Ty1-copia* em *Q. suber*

2.3.1 Digestão de DNA genómico com enzimas de restrição

Digeriram-se 3µg de DNA genómico de *Q. suber* com 50U de *MseI* (Biolabs) a 37°C, em 10µl de tampão de restrição apropriado (NEBuffer 2) e na presença de 1µl BSA (100 µg/ml) para um volume final de 100µl. Para otimizar o tamanho dos fragmentos obtidos foram retiradas alíquotas de 20µl aos 5, 10, 15, 20 e 40min de digestão e, com o objectivo de parar a reacção enzimática, incubou-se cada amostra a 80°C durante 10min. Visualizaram-se 3µl

de acordo com o ponto 2.2 e escolheu-se o tempo de digestão no qual se obtiveram fragmentos com um valor aproximado de 1000bp.

2.3.2 Purificação de DNA

Após a digestão enzimática os fragmentos de DNA foram purificados pelo método de fenol/clorofórmio. Adicionou-se fenol na proporção de 1:1 (v/v) a cada amostra e centrifugou-se brevemente para separar as fracções orgânica e aquosa. Em seguida recolheu-se a fase aquosa (superior e que contém o DNA) para um novo tubo. Adicionou-se um volume de clorofórmio, repetindo a centrifugação anterior e transferiu-se novamente a fase aquosa para um novo tubo. A esta solução adicionaram-se 2 volumes de etanol a 95% previamente arrefecido, 1/10 do volume inicial de NaOAc 3M (pH 4,5) e 20µg de glicogénio. Centrifugou-se a 16.000 g durante 10min, descartou-se o sobrenadante e lavou-se o “pellet” em 100µl de etanol a 80% previamente arrefecido. Centrifugou-se novamente durante 5min para remover algum etanol residual e deixou-se secar o “pellet” ao ar. A ressuspensão do DNA foi feita em 20µl de água destilada e autoclavada.

2.3.3 Preparação do “template DNA” - ligação de adaptadores

Incubaram-se 4pmol de cada adaptador em cadeia simples durante 1h, à temperatura ambiente de forma a obter adaptadores em cadeia dupla (sequência dos adaptadores em cadeia simples: MseI: 5'-GACGATGGATCCTGAG-3' e MseII: 5'-TACTCAGGATCCAT-3'). Posteriormente ligaram-se 2pmol destes adaptadores a 250ng do DNA purificado com 3U de T4 DNA ligase (Promega), 10µl de tampão da enzima e perfazendo com água para 50µl de volume final. A reacção ocorreu durante 2h a 20°C.

2.3.4 Amplificação de DNA por PCR

A estratégia seguida foi a descrita por Pearce (Pearce et al. 1999). Para amplificação da zona de interesse (entre o 1º motivo conservado da *RNaseH* (RTKHID/E) e a LTR 3' (Figura 2.1) utilizaram-se 50ng de DNA ao qual foram ligados os adaptadores e 50pmoles de cada um dos seguintes “primers”: RnaseHI (5'-MGNACNAARCAYATHGA-3') e “primer” do adaptador de *MseI* (5'-GATGGATCCTGAATAA-3'). Para além destes, as misturas de reacção continham 5µl de 10x tampão de PCR (200mM Tris-HCl (pH 8,4), 500mM KCl), 1,5mM de MgCl₂, 0,4mM de dNTP's e 5U de Taq DNA polimerase (Invitrogen) num volume final de 50µl. O programa de PCR consistiu em 2min de incubação inicial a 94°C, seguidos de 30 ciclos de 1min a 94°C, 1,30min a 42°C, 1,30min a 72°C e uma extensão final de 4min a 72°C. Uma pequena quantidade dos produtos obtidos (5µl) foi visualizada em gel de agarose a 1% para confirmar se a amplificação fora eficiente e se os produtos tinham o

tamanho esperado (~1Kb). Em seguida procedeu-se à sua purificação utilizando o “High Pure PCR Product Purification kit” (Roche), de acordo com o protocolo que o acompanha.

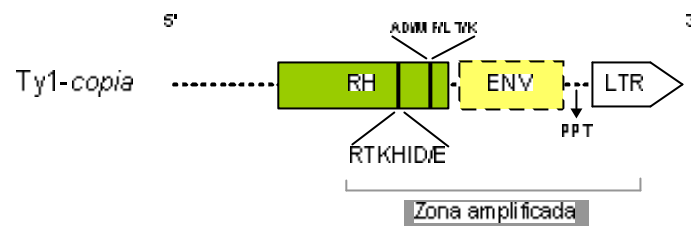


Figura 2.1- Esquema representativo da zona de DNA amplificada em retrotransposições *Ty1-copia* seguindo a estratégia de Pearce *et al.* (1999). É utilizado um “primer” desenhado para o primeiro motivo conservado da *RnaseH* (RTKHID/E) e um “primer” específico para o adaptador ligado ao DNA, na proximidade ou mesmo dentro da LTR 3’ (não representado). A amplificação entre estes “primers” leva ao isolamento de sequências semelhantes à “envelope” existente nos retrovírus, caso estejam presentes. RH – *RnaseH*; ENV – região que codifica para envelope; PPT – região rica em polipurinas (guanina e adenina)

2.4 Isolamento de *RVT*'s de retrotransposições *Ty1-copia* em *Q. suber*

2.4.1 Amplificação de DNA por PCR

Para se proceder ao isolamento de transcriptases reversas (*RVT*'s) de retrotransposições *Ty1-copia* recorreu-se à amplificação de DNA por PCR utilizando primers desenhados para os domínios conservados nesta enzima: 5'-QMDVKTA-3' e 5'-LYVDDM-3' (Figura 2.2). As misturas de reacção continham 100ng de DNA genómico, 25pmoles de cada primer, 2,5µl de 10x tampão de PCR (200mM Tris-HCl (pH 8,4), 500mM KCl), 1,5mM de MgCl₂, 0,4mM de dNTP's e 2,5U de Taq DNA polimerase (Invitrogen) num volume de 25µl. O programa de PCR consistiu em 1min de incubação inicial a 94°C seguido de 30 ciclos de 1min a 94°C, 1,30min a 40°C, 2min a 72°C e uma extensão final de 10min a 72°C.

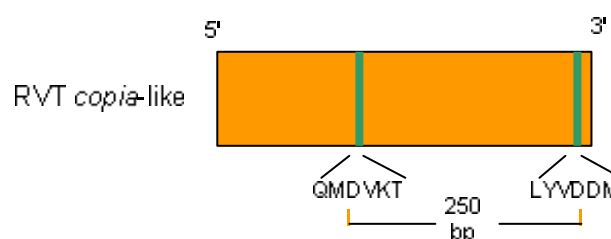


Figura 2.2- Representação esquemática da região que codifica para a transcriptase reversa em retrotransposições *Ty1-copia*. Os “primers” foram desenhados para dois motivos conservados, QMDVKT e LYVDDM.

2.4.2 Isolamento de fragmento(s) de DNA a partir de um gel de agarose

O DNA obtido foi corrido em gel de agarose de acordo com o ponto 2.2. Após a separação dos fragmentos excisaram-se do gel as bandas que continham os fragmentos de DNA de

interesse e eluíram-se estes utilizando o “kit” High Pure PCR Product Purification (Roche, cat. No. 11732676001) seguindo o protocolo que o acompanha.

2.5 Clonagem e sequenciação

2.5.1 Preparação de células competentes (*E. coli* estirpes DH5a e XL1B)

Incubou-se um stock de células em meio 2YT (1,6% bacto-triptona, 1% extracto de levedura, 0,5% NaCl) a 37°C, com agitação, durante a noite. Posteriormente adicionou-se DMSO para uma concentração final de 7%, imediatamente antes de congelar as células a -70°C. Inocularam-se 10µl de células em 250ml de meio SOB (2% bacto-triptona, 0,5% extracto de levedura, 0,01M NaCl, 2,5mM KCl, 0,01M MgCl₂, 0,01M MgSO₄) num frasco de 2L e colocaram-se as culturas a 24°C e com agitação, até atingirem uma densidade óptica de 0,60D a λ=600 nm. Colocou-se o frasco em gelo durante 10min e posteriormente transferiu-se a cultura para um tubo de centrifuga. A centrifugação decorreu a 4°C e 800g (3000 rpm) durante 10min. A ressuspensão do “pellet” foi feita em 80ml de TB [0,01 PIPES (1, 4 Piperazinebis (ácido etanesulfónico)], 0,015M CaCl₂, 0,25M KCl (pH 6,7), 55mM MnCl₂) previamente arrefecido, colocaram-se as células em gelo durante 10min e centrifugou-se novamente nas condições acima referidas. Ressuspendeu-se novamente o “pellet” em 20ml de TB e adicionou-se DMSO para uma concentração final de 7%. Colocaram-se as células novamente em gelo durante 10min e retiraram-se alíquotas de 200 e 400µl, congelando-as imediatamente.

2.5.2 Preparação do “insert”- ligação de produtos de PCR

Ligaram-se 50ng de fragmentos de DNA correspondentes às LTR's 3', isolados e purificados a 50ng de vector pGEM T-easy (Promega) com 3U de T4 DNA ligase em 5µl de 2x tampão (60mM Tris-HCl (pH 7,8), 20mM MgCl₂, 20mM DTT, 2mM ATP 10% PEG) e água, para um volume final de 10µl. A reacção decorreu a 4°C, durante a noite. Relativamente às sequências de *RVT*'s obtidas anteriormente, as condições de ligação foram as mesmas à excepção da quantidade do fragmento, que neste caso foi de 75ng para 50ng de vector.

2.5.3 Transformação

A 200µl de células competentes (*E. coli* estirpe XL1B) adicionaram-se 10ng de vector com o fragmento de interesse clonado. Colocaram-se as células em gelo durante 30min, seguindo-se um choque térmico a 42°C durante 45seg, voltando a colocar as células no gelo durante 2min. Posteriormente adicionaram-se 500µl de meio SOC (2% triptona, 0,5% extracto de levedura, 10mM NaCl, 2,5mM KCl, 10mM MgCl₂, 10mM MgSO₄, 20mM glucose) e incubaram-se as células a 37°C durante 1,30h, com agitação, para permitir a expressão do

gene de resistência à ampicilina. Plaquearam-se as células em meio LA (1L de meio contém 5g de extracto de levedura, 10g de triptona, 5g de NaCl, 16g de agar) contendo IPTG (0,1mg/ml), X-gal (40µg/ml) e ampicilina (0,1mg/ml) e incubaram-se a 37°C, durante a noite.

2.5.4 Selecção das colónias transformadas por PCR

Na transformação obtiveram-se colónias bacterianas azuis (não transformadas) e colónias brancas (transformadas), tendo sido seleccionadas algumas colónias brancas para confirmar a eficiência da transformação. Recolheu-se uma pequena fracção de cada colónia (técnica do palito estéril) e colocou-se num tubo de PCR (PCR cloning), ao qual se adicionou posteriormente 2,5µl de tampão de PCR (200mM Tris-HCl (pH 8,4), 500mM KCl), 1,5mM de MgCl₂, 0,4mM de dNTP's, 25pmoles de cada primer, 2,5U de Taq DNA polimerase (Invitrogen, cat. No. 18038-026) e água para um volume final de 25µl. Os primers utilizados têm a seguinte sequência: M13 (Fw): 5'- GTAAAACGACGGCCAGTG -3' e M13 (Rv): 5'- GGAAACAGCTATGACCATG -3'. O programa de PCR consistiu em 5min a 96°C seguido de 30 ciclos de 45seg a 94°C, 45 seg a 55°C, 1,30min a 72°C e 5min finais de extensão também a 72°C. Correram-se 15µl dos produtos de PCR como referido no ponto 2.2 e seleccionaram-se as colónias transformadas. Estas foram crescidas em meio LB (1L de meio contém 5g de extracto de levedura, 10g de triptona, 5g de NaCl) com ampicilina (0,1 mg/ml) e incubou-se a 37°C durante a noite, com agitação.

2.5.5 Extracção de DNA plasmídico e sequenciação

Extraíu-se DNA plasmídico com o Qiagen Spin Miniprep Kit (Qiagen) utilizando o protocolo que acompanha o "kit" e solicitou-se a sequenciação das amostras à empresa Macrogen (Coreia do Sul).

3. Análise bioinformática

As sequências obtidas foram analisadas relativamente a homologias com outras sequências recorrendo à base de dados *GenBank*, pela utilização do algoritmo BLAST. Para confirmação da possível existência de domínios transmembranares nas sequências isoladas entre os domínios da *RnaseH* e a LTR 3' de retrotransposões Ty1-*copia* recorreu-se à utilização do software TMpred (Hofmann and Stoffel 1993). O alinhamento de sequências e respectivos dendogramas foram realizados pelo método neighbor joining, utilizando o software Bionumerics, versão 3.5.

4. Análise Citogenética Molecular

4.1 Silanização de lâminas

As lâminas de vidro foram lavadas durante 30min em água quente com detergente removendo possíveis resíduos de gordura. Em seguida mergulharam-se sequencialmente durante 10seg numa solução de 2% de silano (Trimetoxysilano) em acetona, em acetona pura e finalmente em água destilada.

4.2 Pré-tratamento do material biológico

O pré-tratamento do material a utilizar é de extrema importância, uma vez que o prepara para os procedimentos experimentais a que vai ser sujeito. Neste caso, o tratamento G-mitótico é crucial para a obtenção de um maior número de metafases e para a individualização dos cromossomas, permitindo uma melhor visualização da sua morfologia. Por outro lado, a fixação do material permite preservar a sua estrutura, indispensável para a análise subsequente.. Como tal, após a germinação das bolotas de *Q. suber* excisaram-se os ápices radiculares com cerca de 1mm que foram seguidamente colocados numa solução de a-bromonaftaleno durante 3,30h, no escuro de modo a obter c-metafases (Sharma and Sharma 1999). Os ápices radiculares foram seguidamente lavados em água destilada e fixados em etanol/ácido acético (3:1) à temperatura ambiente, durante a noite. No dia seguinte renovou-se o fixador e colocou-se o material a -20°C até a sua utilização (Schwarzacher and Heslop-Harrison 2000).

4.3 Técnica de drop em raízes de *Quercus suber*

O material fixado foi lavado duas vezes em 1xEB (diluição 1:10 da solução “stock” 10xEB: 40ml de ácido cítrico monohidratado 0,1M e 60ml de citrato trissódico desidratado 0,1M; pH 4,8), 10min cada lavagem com agitação. Os ápices radiculares foram isolados e depois da coifa removida com o auxílio de uma lupa binocular foram digeridos enzimaticamente com o objectivo de remover a parede celular.. A mistura enzimática continha 2% (p/v) celulase, 1,5% macerozima, 0,3% pectoliase Y23, 0,03% EDTA em 2XSSC, pH 4,2 (Zoldos et al. 1999) e a digestão decorreu a 37°C durante 4h. O “pellet” obtido foi centrifugado a 960g durante 4min, removeu-se o sobrenadante e adicionaram-se 400µl de 1xEB. Repetiu-se este passo 3 vezes e após a última centrifugação removeu-se o sobrenadante e adicionaram-se 350µl de 3:1 ao “pellet” formado. Agitou-se ligeiramente a mistura no tubo eppendorf durante 30seg e colocou-se uma gota da mistura numa lâmina, observando-se em seguida ao microscópio óptico de contraste de fase. Como os núcleos se encontravam maioritariamente envolvidos por uma bainha citoplasmática, o material vegetal foi sujeito à acção de ácido acético a 60% durante 1,30min. Centrifugou-se como referido acima, removeu-se o

sobrenadante e fixou-se novamente o material em 3:1. Observou-se uma gota do material novamente ao microscópio e repetiu-se o passo do ácido acético até se obterem núcleos sem citoplasma. Quando se atingiu esta fase, fixou-se o material em 3:1 e colocaram-se 2 gotas da mistura em cada lâmina, deixando-se secar ao ar. As lâminas assim preparadas foram guardadas a 4°C até à sua utilização.

5. Análise de DNA pelo método de C-PRINS (Cycling primed *in situ* DNA labeling)

Com o objectivo de estudar a localização das sequências isoladas neste trabalho no genoma de *Q. suber*, nomeadamente *RVT*'s de retrotransposões *Ty1-copia* e sequências “*env-like*”, seleccionou-se uma sequência de cada um dos tipos para ser utilizada como sonda. No entanto o tamanho das sequências, cerca de 250bp para a *RVT* e 480bp para a sequência “*env-like*” colocou problemas na sua detecção pela técnica de FISH (Fluorescent *in situ* hybridization). Como tal optou-se pela técnica de PRINS, mais propriamente C-PRINS, que permite a visualização de genes de cópia única e de reduzida dimensão (Cinti et al. 1993; Kadandale et al. 2000; Wachtel and Tharapel 2006). Foi então desenhado um par de “primers” específico para cada sequência escolhida com o software PrimerPremier (versão 5.0) e adaptou-se a metodologia descrita por Mukai e Appels (1996) ao material utilizado. Adicionalmente foi também testada a distribuição da *RVT* de um retroelemento *Ty3-gypsy* previamente isolado em sobreiro e cujo produto da sua amplificação é 623bp.

5.1 Pré-tratamento do material

As lâminas previamente preparadas foram incubadas numa solução de 70% de formamida em 2xSSC a 70°C durante 2min, com agitação ligeira de modo a desnaturar o DNA, seguindo-se uma desidratação durante 5min em 70% de etanol a -20°C e finalmente em 90% de etanol à temperatura ambiente, durante 5 min.

5.2 Preparação e adição da mistura de PCR

Delimitou-se a área com o material biológico em cada preparação pela colocação de uma “geneframe” (janela adesiva) e aplicou-se 30µl da mistura de PCR com a seguinte composição: 60pmol de cada primer, 10mM Tris-HCl (pH 8,3), 50mM de KCl, 200µM de dNTP, 60µM de dUTP-16-biotina (Roche) ou dUTP-11-digoxigenina (Roche) 3mM de MgCl₂, 0,8µg/µl de BSA, 3U de Taq e água de modo a perfazer o volume final (30µl). Nas preparações usadas como foi aplicada a mesma mistura de PCR à excepção da enzima Taq, que não foi adicionada. Após a adição da mistura, selou-se cada janela com lamela

adequada lavada em álcool e evitando a formação de bolhas de ar. A sequência dos primers usados é a seguinte:

RVT Ty3- gypsy: 5'-AATTCAACCGAGTCAAAGTTCC -3' (Fw)

5'- CAACCCAAGAAATCCTCGCAG -3' (Rv)

RVT Ty1- copia: 5'-TAAATGGGGACCTAGAGGAAGAGAT-3' (Fw)

5'-TCGACGTATAGGCTAATAATAACAT-3' (Rv)

“*Envelope-like*”: 5'- AGTGTGTCTGTAGTTTTTAGCC-3' (Fw)

5'- TTCCTAAAGCTGTGCTGTAGCCTGC-3' (Rv)

5.3 C – PRINS (Cycling Primed *in situ* DNA Labeling)

Cada preparação foi colocada dentro de uma câmara plástica contendo óleo mineral, para evitar que a mistura de reacção secasse durante o tempo de amplificação e, por outro lado, mantendo a temperatura estável ao longo de toda a lâmina. Selou-se cada câmara, evitando a formação de bolhas de ar e colocaram-se as preparações num termociclador. O programa de amplificação para as sequências de *RVT*s do tipo *Ty3-gypsy* consistiu em 4min a 57°C seguidos de 10min a 72°C. Seguiram-se 30 ciclos de 94°C durante 2min, 2min a 57°C e 8min a 72°C. Para as sequências de *RVT*s de *Ty1-copia* e “*envelope-like*” o programa foi exactamente o mesmo, exceptuando na temperatura de ligação dos primers ao DNA, 50°C. Todos os passos desta técnica são mais demorados relativamente a uma normal reacção de PCR com o DNA em suspensão, porque o material genético está fixado e o acesso da *Taq* polimerase ao mesmo pode estar de alguma forma dificultado.

5.4 Lavagens pós-reacção

Retiraram-se as lâminas do termociclador e lavaram-se em clorofórmio para retirar a gordura do óleo mineral. Posteriormente retirou-se a lamela de cada lâmina e seguiram-se as lavagens. A primeira lavagem foi feita em 70% de etanol em 0,05% SDS, a segunda em 2xSSC e a terceira em 4xSSC, sempre com agitação e durante 5min cada.

5.5 Detecção dos locais onde ocorreu amplificação

A detecção dos padrões de distribuição das sequências amplificadas foi feita pela incubação do material com um anticorpo associado a um fluorocromo. Para a detecção da biotina ligada ao nucleótido dUTP usado para amplificar as sequências de *RVT* do tipo *Ty3-gypsy* foi utilizado o conjugado streptavidina associado com o fluorocromo Cy3 (Sigma) numa diluição de 1:200 numa solução de 4xSSC contendo 3% de BSA (albumina de soro bovino). O fluorocromo Cy3 possui um pico de absorção nos 554nm (verde) e emite a 568-574nm (vermelho). Para as sequências de *RVT* do tipo *Ty1-copia* e para as “*envelope-like*” utilizou-se o anticorpo anti-digoxigenina associado com o fluorocromo FITC (Roche) numa diluição

de 1:20 na mesma solução. Os picos de excitação e de emissão do fluorocromo FITC ocorrem aos 495nm (azul) e 520nm (verde-amarelado), respectivamente (Schwarzacher and Heslop-Harrison 2000). A incubação foi feita em câmara húmida a 37°C, no escuro e durante 1h. Posteriormente lavaram-se as lâminas nas seguintes soluções: 4xSSC (1min), 0,2% Tween-20 em 4xSSC (10min), 4xSSC (10min) e 2xSSC (5min), todas efectuadas no escuro. Montaram-se as lâminas em 2 µg/ml de DAPI em citifluor e após 3h de incubação no escuro observou-se ao microscópio de epifluorescência. O DAPI é um corante fluorescente de DNA que absorve luz UV ($\lambda=385\text{nm}$) e fluoresce no azul ($\lambda=461\text{nm}$) marcando regiões ricas em A-T (Kapuscinski 1995).

5.6 Visualização, captação e processamento de imagens

A visualização dos sinais fluorescentes foi feita num microscópio de epifluorescência (Zeiss) e as imagens captadas através de uma câmara fotográfica (Zeiss AxioCamMRc5 4.0) acoplada ao microscópio. O processamento das imagens foi feito utilizando o software Adobe Photoshop (versão 7.0).

6. Hibridação *in situ* fluorescente (FISH)

A hibridação *in situ* permite a identificação de sequências específicas de ácidos nucleicos, DNA ou RNA, em cromossomas, núcleos, células e secções de material conservado (Schwarzacher and Heslop-Harrison 2000). A utilização desta técnica teve como objectivo facilitar a localização cromossómica das sondas usadas em PRINS utilizando uma sonda cuja localização em *Q. suber* é conhecida : pTa71 (rDNA 45S).

6.1 Preparação da sonda

Neste trabalho usou-se como sonda pTa71, que contém genes de DNA ribossomal (18S, 5.8S e 26S) de trigo e que se localiza nas regiões organizadoras do nucléolo (NOR-nucleolar organizer region). Esta foi marcada com digoxigenina pela incorporação de dUTP-11-dig (Roche) pela técnica de “nick translation” (Roche) seguindo o protocolo que acompanha o produto.

6.2 Pré-tratamento do material

Este passo tem como objectivo aumentar a especificidade da ligação da sonda ao DNA, removendo RNA e proteínas às quais a sonda se poderia ligar. Por outro lado, a digestão com pepsina facilita o acesso da sonda ao DNA pela permeabilização dos núcleos. Como tal, o material foi digerido com pepsina, usando uma diluição de 1:75 de uma solução stock (500 µg/ml) em 0,01M de HCl. Adicionaram-se 200µl desta solução a cada lâmina e cobriu-

se com lamela de plástico, incubando a 37°C por 10min em câmara húmida. Posteriormente as lâminas foram lavadas em 2xSSC durante 5min, com agitação, duas vezes. Seguiu-se uma digestão do material com RNase (Sigma) usando uma diluição de 1:100 duma solução stock (100 mg/ml) em 2xSSC. Adicionaram-se 200µl a cada lâmina e cobriu-se com lamela de plástico, sendo incubadas durante 1h, a 37°C em câmara húmida. Em seguida as lâminas foram lavadas em 2xSSC 3 vezes durante 5min cada e com agitação. O material foi depois desidratado, imergindo as lâminas em 70% de etanol durante 3min, com agitação, sendo transferidas depois para etanol a 100% durante 3min, sem agitação e finalmente deixando-as secar ao ar.

6.3 Preparação da mistura de hibridação

Um dos pontos críticos desta técnica consiste na optimização das condições de hibridação que depende, entre outros parâmetros, da estringência. Esta corresponde à percentagem de nucleótidos que hibrida correctamente entre a sonda e a sequência alvo e pode ser alterada modificando a quantidade de formamida e de sais que existe na mistura (Schwarzacher and Heslop-Harrison 2000). Nas hibridações realizadas neste trabalho a estringência utilizada foi 76%, obtida com 40% de formamida e 1xSSC.

Preparou-se a mistura de hibridação, cuja composição é a seguinte: formamida 40%, 1x SSC, 6µl de sulfato de dextrano 50% (p/v), 0,5µl de SDS 10% (p/v), DNA de esperma de salmão (5µg), 200 ng de sonda (pTa71) e água destilada e esterilizada para um volume final de 30µl por lâmina.

6.4 Desnaturação e hibridação

Para que a hibridação ocorra é necessário que quer a sonda quer o DNA alvo se encontrem em cadeia simples e como tal, é necessário desnaturar ambos (Schwarzacher and Heslop-Harrison 2000). Assim, desnaturou-se a mistura de hibridação a 95°C durante 7min e depois colocou-se em gelo durante 5min para evitar uma possível renaturação do DNA. Adicionaram-se 30µl da mistura a cada lâmina e cobriu-se com lamela de plástico, tendo o cuidado de eliminar bolhas de ar que se tivessem formado. Colocaram-se as lâminas no termociclador e procedeu-se à desnaturação do material e sonda da seguinte forma: 85°C, 5min; 75°C, 1min; 65°C, 1min; 55°C, 1min; 45°C, 1min; 37°C, 3min. Retiraram-se as lâminas do aparelho e incubaram-se em câmara húmida a 37°C, durante a noite.

6.5 Lavagens pós-hibridação

Posteriormente à hibridação, as lâminas foram lavadas para remover a mistura de hibridação e porções de sonda que tenham hibridado de forma inespecífica. A primeira

lavagem foi feita em 2xSSC com agitação durante 3min, a 42°C. Seguiram-se duas lavagens numa solução de formamida a 20% em 0,1xSSC, durante 5min e duas lavagens em 2xSSC, 5min cada. A temperatura utilizada foi de 42°C. Fez-se posteriormente uma lavagem à temperatura ambiente em 2xSSC durante 10min e duas lavagens em 4x SSC/Tween (0.2%) durante 5min, à mesma temperatura.

6.6 Detecção dos locais de hibridação

Incubaram-se as lâminas numa solução 4xSSC/Tween (0.2%) com 5% de BSA para bloquear o acesso do anticorpo a proteínas às quais este se poderia ligar e formar falsos positivos. Adicionaram-se 200µl da solução por lâmina, cobrindo-se com lamela de plástico e incubou-se 5min à temperatura ambiente. Em seguida adicionaram-se 50µl de uma solução de anti-digoxigenina FITC numa proporção de 1:20 (Roche,) em 4xSSC/Tween (0.2%) com 5% BSA a cada lâmina e incubaram-se as mesmas novamente a 37°C durante 1h, em câmara húmida, no escuro. Para retirar o excesso de solução e anticorpo lavaram-se as lâminas em 4xSSC/Tween (0.2%) durante 5min, 3 vezes e com agitação. Montaram-se finalmente as lâminas em 2µg/ml de DAPI em citifluor, cobrindo com lamela de vidro e retirando as bolhas de ar que pudessem existir, incubando-as no escuro durante 2h.

6.7 Visualização, captação e processamento de imagens

De acordo com o descrito no ponto 5.6.

Capítulo 3. Resultados

1. Isolamento de seqüências de *RVT*'s de retrotransposões *Ty1-copia* em DNA genômico de *Quercus suber*

Com o objectivo de isolar seqüências de transcritases reversas (*RVT*) de retrotransposões *Ty1-copia* no genoma de *Quercus suber*, recorreu-se à amplificação de DNA por PCR. Para tal desenhou-se um par de “primers” degenerados para motivos conservados neste gene, nomeadamente o domínio QMDVKT e YVDDM (Figura 2.2). A amplificação em DNA genômico de sobreiro gerou um conjunto de fragmentos de diferentes tamanhos moleculares (Figura 3.1-A) para além do fragmento esperado de aproximadamente 300bp resultante da amplificação entre os domínios da *RVT* acima referidos (Matsuoka and Tsunewaki 1999). Adicionalmente obteve-se um fragmento de tamanho um pouco superior, com cerca de 550bp, e um conjunto de fragmentos de maior dimensão a variar entre 1 e 3Kb. Este conjunto de fragmentos acima dos 300bp pode resultar de uma ligação inespecífica dos “primers” a outros locais do genoma, uma vez que estes eram degenerados, relativamente pequenos (20bp) e a temperatura de “annealing” utilizada (40°C), diminui ainda mais a especificidade da ligação dos “primers” ao DNA.

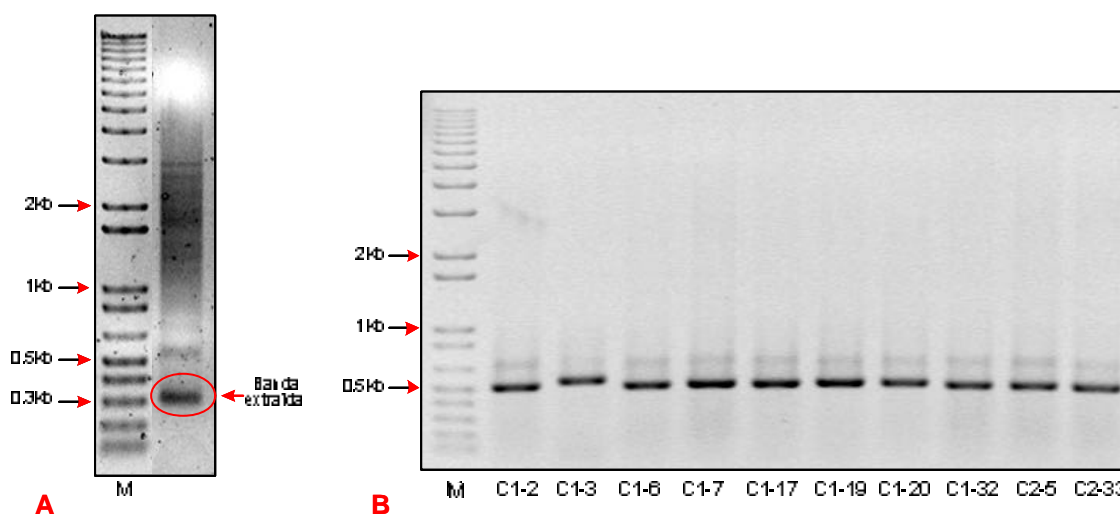


Figura 3.1-A - Amplificação de DNA por PCR de seqüências parciais de transcritases reversas (*RVT*'s) de retrotransposões *Ty1-copia* no genoma de *Quercus suber*. O segmento de DNA com aproximadamente 300bp corresponde ao produto com o tamanho esperado e foi extraído, clonado e sequenciado. M – marcador 1kb plus. B - Na parte inferior da figura estão assinalados os clones de *Q. suber* sequenciados, cuja busca na base de dados *GenBank* revelou homologia com seqüências de *RVT*'s de outros organismos. O tamanho real de cada seqüência é inferior ao apresentado na imagem em cerca de 200bp, que resultam da amplificação duma pequena parte do plasmídeo onde cada seqüência foi clonada.

O fragmento de DNA de 300bp foi extraído do gel, eluído, clonado e posteriormente seleccionados 15 clones que foram sequenciados. As sequências obtidas foram analisadas na base de dados *GenBank* tendo-se verificado que apenas 10 (66,6%) das 15 apresentavam homologias com sequências de transcriptases reversas (Figura 3.1-B). Utilizando as sequências de aminoácidos correspondentes, obtidas pela tradução putativa das sequências de nucleótidos, procedeu-se à análise filogenética pelo método neighbor joining, usando o software Bionumerics (versão 3.5), tendo-se obtido o seguinte dendograma (Figura 3.2).

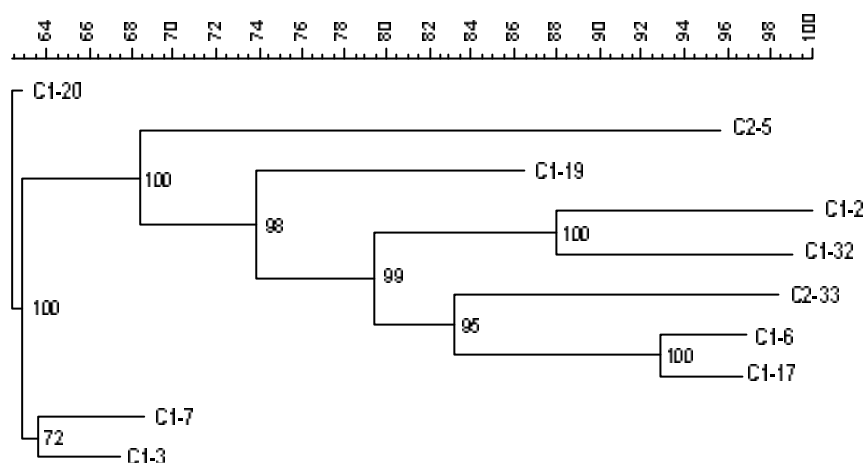


Figura 3.2 – Análise filogenética de sequências de RVT's de retrotransposões Ty1-copia de *Quercus suber*, pelo método neighbor joining (Saitou and Nei 1987). O dendograma obtido resulta de um alinhamento das sequências de aminoácidos utilizando o software Bionumerics (versão 3.5). Os números em cada nó correspondem aos valores percentuais de “bootstrap” relativos a 1000 réplicas, suportando a estabilidade de cada nó.

A análise do dendograma permite inferir que a população de sequências isoladas em *Q. suber* apresenta alguma variedade no que respeita ao grau de homologia, que varia entre 63 e 93%. Como se pode observar existe um único cluster suportado por um “bootstrap” de 100% e que engloba todos os elementos à excepção do clone C1-20, que pode ser considerado como “outgroup”. Este “cluster” pode ainda ser sub-dividido em dois sub-clusters, um formado pelos clones C1-3 e 7, com um bootstrap de 72% e com 64% de homologia, e o outro formado pelos restantes clones com um bootstap de 100% e com 68,5% de homologia. Dentro do “cluster” maior observa-se um grupo de sequências cuja homologia entre si é de aproximadamente 80%, suportada por um valor de “bootstrap” de 99%.

Com o objectivo de determinar a relação entre as sequências de transcriptases reversas identificadas em sobreiro com sequências de RVT's de retrotransposões “copia-like” de

outras espécies vegetais, fez-se uma outra análise filogenética pelo método neighbor joining (Saitou and Nei 1987) e o dendograma obtido está representado na figura 3.3.

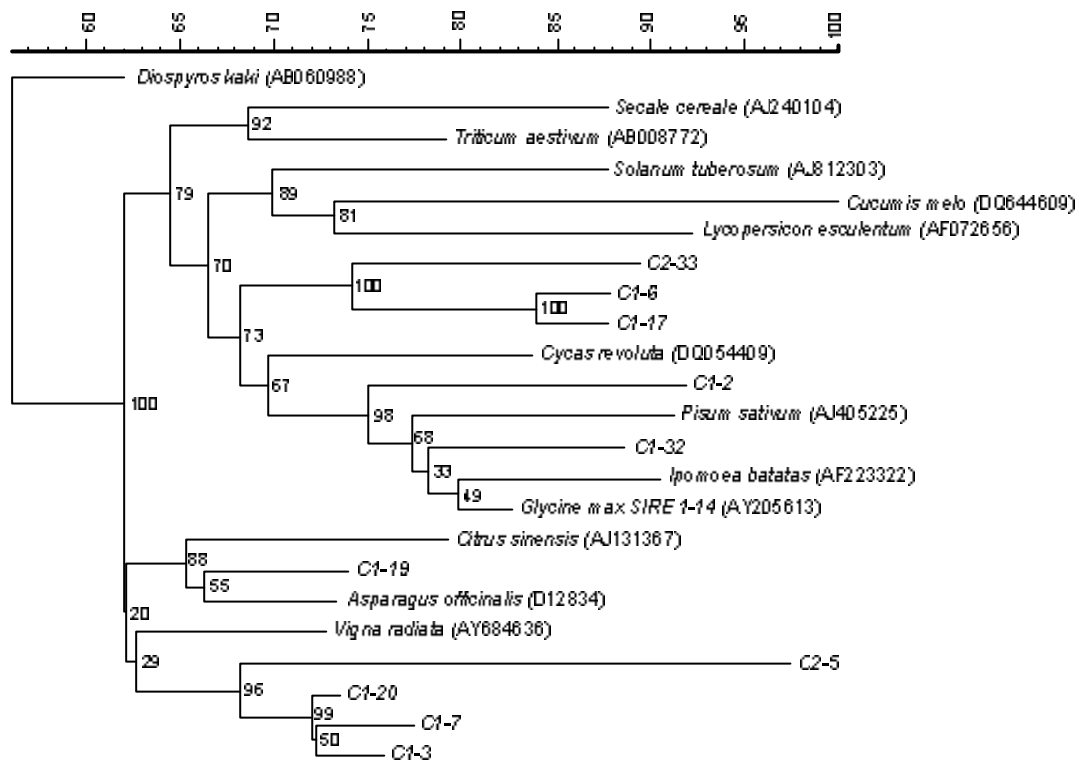


Figura 3.3 - Análise filogenética de sequências de transcritases reversas (RVT) de retrotransposições *Ty1-copia* pelo método Neighbor joining (Saitou and Nei 1987). O dendograma obtido resulta de um alinhamento das sequências de aminoácidos isoladas em *Q. suber* com sequências previamente identificadas noutras espécies, utilizando o software Bionumerics (versão 3.5). Os 10 clones isolados em *Q. suber* designam-se C1-2, 3, 6, 7, 17, 19, 20, 32, C2-5 e 33. Os números em cada nó correspondem aos valores percentuais de “bootstrap” relativos a 1000 réplicas e os números de acesso ao *GenBank* estão associados a cada sequência.

A análise do dendograma obtido permite verificar que as sequências estão agrupadas em dois grandes clusters, exceptuando a sequência obtida de *Diospyros kaki* que constitui um “outgroup”, partilhando apenas 56% de homologia com as restantes. Nos mesmos clusters onde se encontram as sequências de sobreiro encontram-se também sequências de *RVT*s isoladas de espécies vegetais pertencentes a duas divisões taxonómicas: Cycadophyta e Magnoliophyta com nove famílias representadas e treze géneros (incluindo o género *Quercus*). O primeiro grande “cluster”, suportado por um “bootstrap” de 79%, agrupa sequências de sobreiro com sequências provenientes de famílias distintas que envolvem divisões taxonómicas: Cycadophyta (*Cycas revoluta*) e Magnoliophyta (restantes espécies). O segundo “cluster” agrupa sequências de sobreiro e sequências de três famílias distintas: Rutaceae (*Citrus cinensis*), Asparagaceae (*Asparagus officinalis*) e Fabaceae (*Vigna*

radiata). Dentro deste o clone C1-19 partilha uma homologia superior com as sequências de *Citrus cinensis* e *Asparagus officinalis* que com os 4 clones de sobreiro que restam, formando estes um grupo com 74% de homologia entre si e suportado por 96% de “bootstrap”.

2. Padrões de distribuição de *RVT*s de retrotransposões *gypsy* e *copia*-like no genoma de *Quercus suber* por C-PRINS (PCR *in situ*)

Com o objectivo de determinar os padrões de distribuição de *RVT*s do tipo *Ty1-copia* no genoma de sobreiro seleccionou-se a sequência C1-3 obtida anteriormente, para ser utilizada como sonda. A sequência foi seleccionada com base no seu tamanho e grau de homologia com sequências de *RVT*s de outros organismos. Adicionalmente, foi usada como sonda uma sequência correspondente à *RVT* de um outro retroelemento isolado em *Q. suber* pertencente à classe *Ty3-gypsy*, anteriormente caracterizado no Laboratório de Genética do ISA. O tamanho de ambas as sequências, (244bp para a sequência *copia* e 650bp para a sequência *gypsy*), colocou alguns problemas na sua detecção pela técnica de FISH pelo que se optou pela técnica de PCR *in situ*. Desenhou-se um par de “primers” específico para cada sequência e com o objectivo de testar a eficiência dos “primers” efectuou-se uma amplificação por PCR em DNA genómico. A amplificação foi bem sucedida, tendo gerado um produto único e com o tamanho esperado, aproximadamente 250bp (figura 3.4).

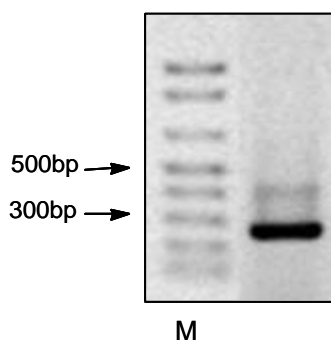


Figura 3.4- Amplificação da sequência “*copia*-like” seleccionada em DNA genómico de *Q. suber* pela técnica de PCR. A coluna M corresponde ao marcador 1Kb plus. A amplificação gerou um produto com o tamanho esperado, 250bp, confirmando a especificidade dos “primers” utilizados.

A amplificação *in situ* das sequências de *RVT*s *gypsy* e *copia*-like foi feita em cromossomas e núcleos de sobreiro isolados de ápices radiculares de *Q. suber*. Relativamente à *RVT* pertencente a retroelementos *Ty3-gypsy* obteve-se uma marcação dispersa e com aspecto “pulverizado” por todo o genoma, quer em núcleos interfásicos (figura 3.6 B e D) quer em cromossomas metafásicos (figura 3.5 B, D e figura 3.6 B e D). No entanto observa-se maior intensidade de marcação nalgumas zonas perinucleolares em

interfase e na extremidade de alguns cromossomas. Estes locais de maior marcação correspondem ao *locus* major de rDNA 45S, como se provou com a hibridação *in situ* posteriormente efectuada com a sonda de trigo pTa71 (figura 3.5 B e figura 3.6 B). A sobreposição das duas sondas confirmou a co-localização dos sinais obtidos para a *RVT* com os NOR1 (figura 3.5 D e figura 3.6 D). Em cromossomas menos contraídos é possível observar uma forte marcação da sonda da *RVT* na região centromérica de um cromossoma (figura 3.5-B e figura 3.6 D).

A técnica de C-PRINS efectuada para a detecção da sequência de *RVT copia*-like evidenciou um padrão de distribuição de marcação diferente comparativamente à sequência *gypsy*-like. Nos núcleos interfásicos a marcação obtida com esta sonda é uniforme não sendo detectada nenhuma região de marcação mais intensa (figura 3.7).

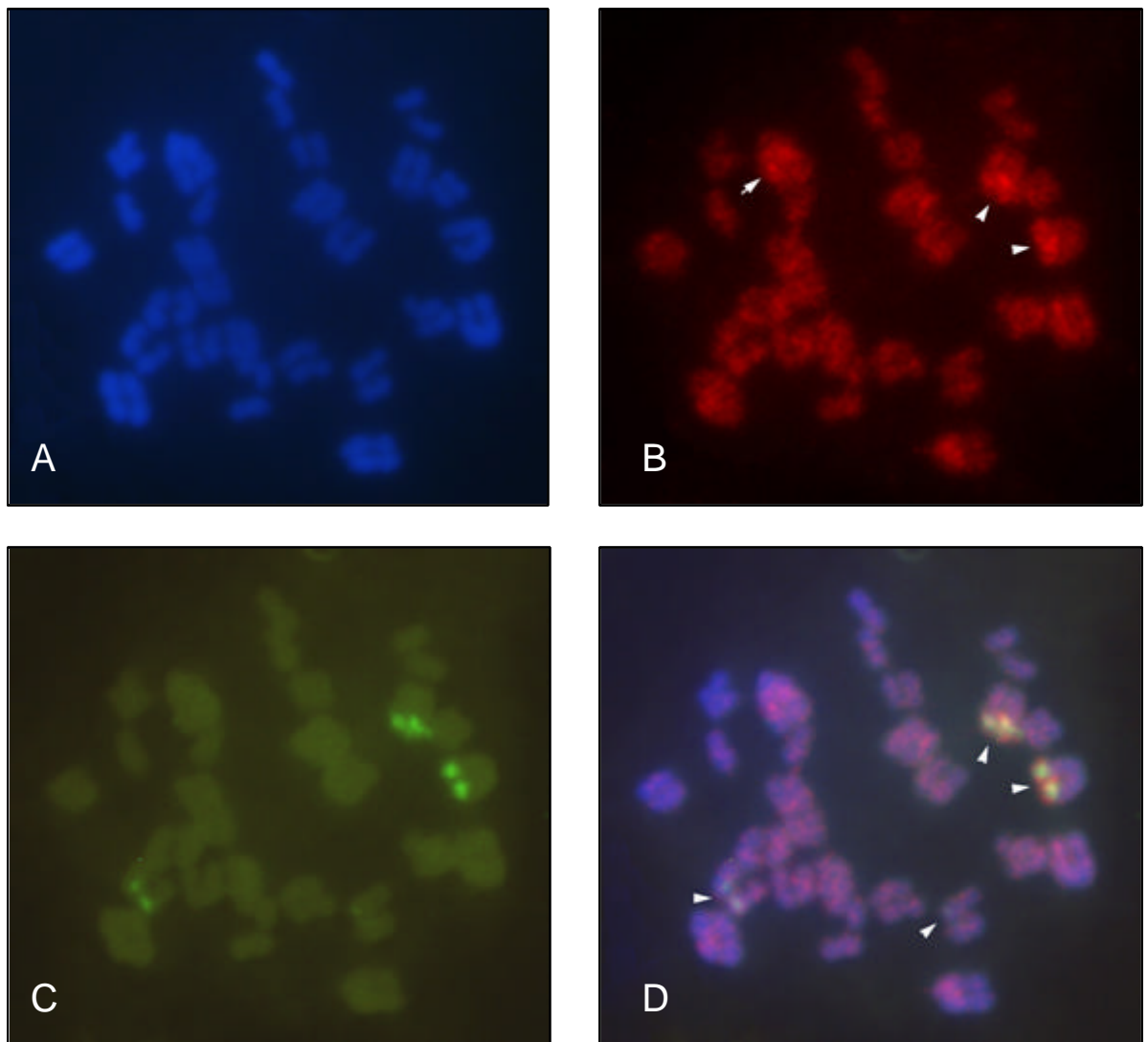


Figura 3.5 – Amplificação da *RVT* de um retrotransposão *Ty3-gypsy* (vermelho) por G-PRINS em cromossomas isolados de ápices radiculares de *Q. suber*. A *RVT* exibe uma marcação dispersa pelo genoma, sendo no entanto mais forte na extremidade de alguns cromossomas (B-setas à direita). Estes locais correspondem ao *locus* major de rDNA 45S, NOR1 (verde), confirmado pela hibridação *in situ* efectuada posteriormente com a sonda de trigo pTa71 (C). A sobreposição das duas sondas confirmou a co-localização dos sinais obtidos para a *RVT* com os NOR1 (D-setas à direita) mas não com o NOR2 (D-seta à esquerda e seta central). Em cromossomas menos contraídos é possível observar uma forte marcação da sonda da *RVT* na região centromérica de um cromossoma (B-seta à esquerda).

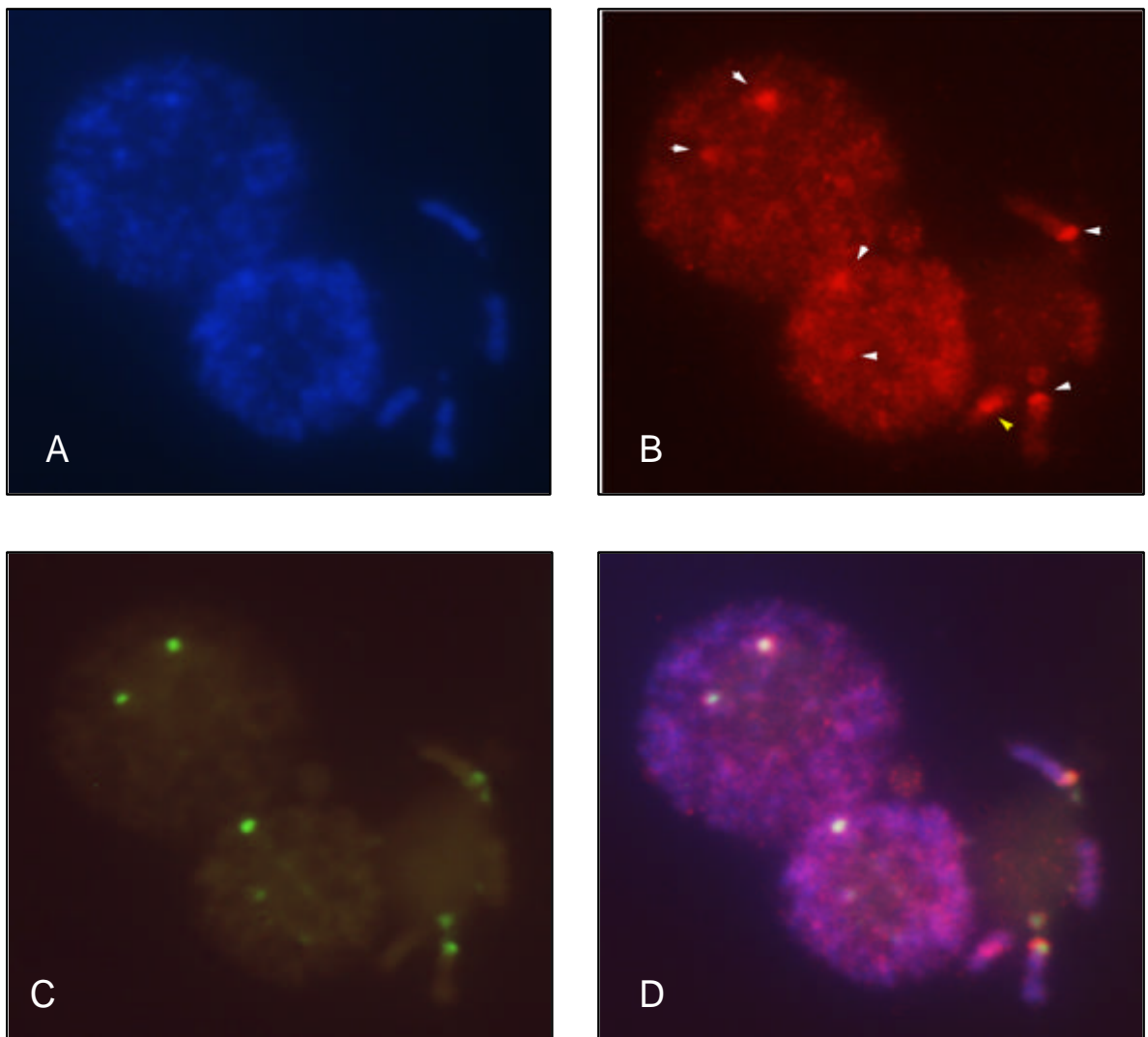


Figura 3.6 - Amplificação da *RVT* de um retrotransposão *Ty3-gypsy* (vermelho) por C-PRINS em núcleos e cromossomas isolados de ápices radiculares de *Q. suber*. A *RVT* exibe uma marcação dispersa e com aspecto “pulverizado” por todo o genoma, quer em núcleos interfásicos quer em cromossomas metafásicos (B). No entanto observa-se maior intensidade de marcação nalgumas zonas perinucleares em interfase e na extremidade de alguns cromossomas (B-setas brancas). Estes locais correspondem ao *locus* major de rDNA 45S, NOR1 (verde), confirmado pela hibridação *in situ* efectuada posteriormente com a sonda de trigo pTa71 (C). A sobreposição das duas sondas confirmou a co-localização dos sinais obtidos para a *RVT* com os NOR1 (D). Em cromossomas menos contraídos é possível observar uma forte marcação da sonda da *RVT* na região centromérica de um cromossoma (B-seta amarela).

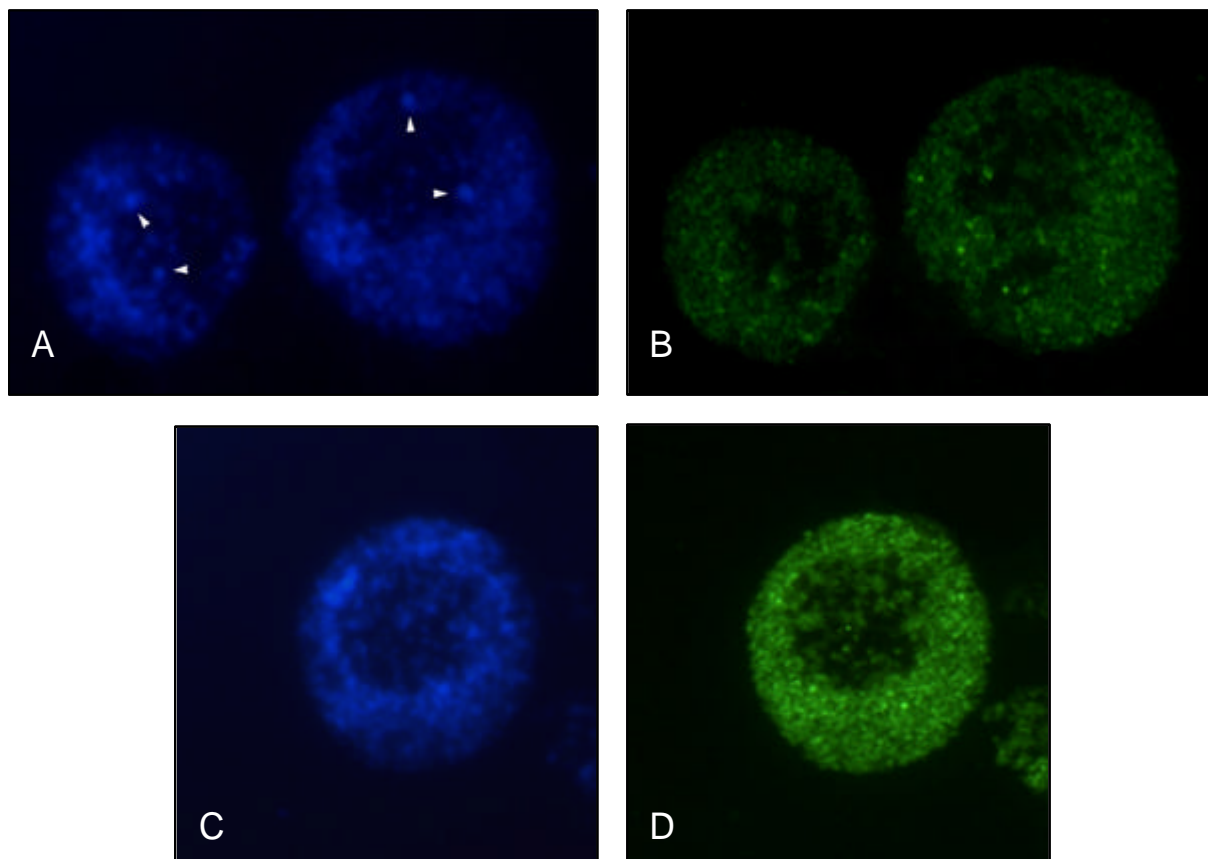


Figura 3.7 – Amplificação de uma sequência *RVT* de retrotransposões *Ty1-copia* (verde) por C-PRINS em núcleos isolados de ápices radiculares de *Q. suber*. A marcação obtida com esta sonda é uniforme pelo genoma (B e D), de aspecto ponteadado, não existindo uma marcação particular nalguma região, por exemplo, nos *loci* de rDNA 45S, NOR1 (A-setas), visíveis como blocos de coloração mais intensa com DAPI (azul).

3. Isolamento de sequências “*env-like*” em retrotransposões *Ty1-copia* de *Quercus suber*

Com o objectivo de isolar sequências do tipo “*env-like*” em retrotransposões *Ty1-copia* de sobreiro seguiu-se a estratégia descrita por Pearce (Pearce et al. 1999). No caso desta sequência estar presente na estrutura do retrotransposão a sua localização será entre o gene que codifica para a *RNaseH* e a LTR 3’ (Figura 3.8). O isolamento de sequências *RNaseH*-LTR foi feito por PCR, utilizando um “primer” degenerado para um dos motivos conservados da *RNaseH* (RTKHID/E) e outro “primer”, específico para um adaptador, que foi ligado previamente aos fragmentos de DNA seleccionados para a amplificação. Esta estratégia gerou um conjunto de fragmentos de DNA cuja dimensão variava entre os 400bp e 1Kb, que foram purificados, clonados e sequenciados num total de 18 clones (Figura 3.9).

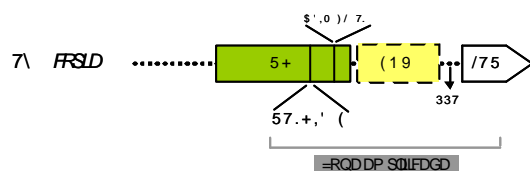


Figura 3.8 - Esquema representativo da zona de amplificação em retrotransposões *Ty1-copia*, em DNA genómico de *Q. suber*. Um dos “primers” foi desenhado para o 1º motivo conservado da *RNaseH*, RTKHID/E, e o outro é específico de um adaptador que foi ligado ao DNA na proximidade ou mesmo dentro da LTR 3’. Caso estejam presentes, serão isoladas sequências “*env-like*”.

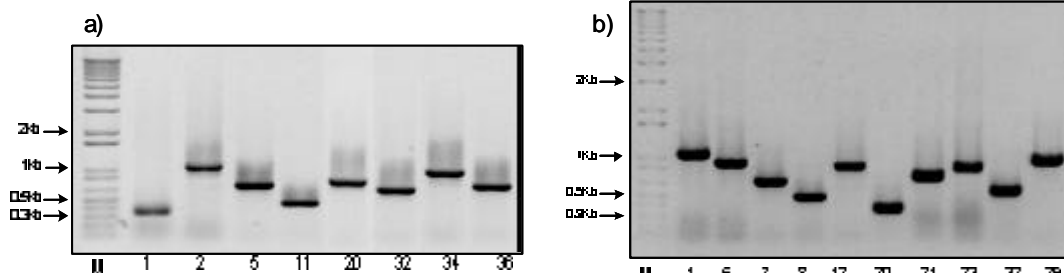


Figura 3.9 – Amplificação de DNA por PCR de sequências entre a *RNaseH* e a LTR 3’ de retrotransposões *Ty1-copia* no genoma de *Q. suber*. a) Clones B2-1, 2, 5, 11, 20, 32, 34 e 36. b) Clones B3-1, 6, 7, 8, 12, 21, 23, 29 e 32. Em ambas as imagens a coluna M corresponde ao marcador de tamanho molecular (1Kb plus). À excepção de quatro clones (B2-1, B3-7, 20, 21) todos os restantes apresentam homologia com sequências “*env-like*” por pesquisa na base de dados *GenBank* ou utilizando o software TMpred, que analisa as sequências e prevê a existência de domínios transmembranares e a sua orientação na membrana celular.

Analisando as sequências com o software TMpred (Hofmann and Stoffel 1993), que prevê a existência de domínios transmembranares e determina a sua orientação na membrana celular, e fazendo uma pesquisa relativamente a homologias com outras

sequências na base de dados *GenBank*, verificou-se que catorze das dezassete sequências isoladas (82%) apresentavam a possibilidade de conter sequências do tipo “*env-like*”. Destas catorze, cinco apresentam homologia a nível dos aminoácidos com sequências do tipo “*envelope*” de outros organismos, nomeadamente *HIV-1* e *SRV-2* (Simian retrovirus). Por outro lado, doze das catorze sequências analisadas pelo TMpred exibiram valores superiores a 500 nos gráficos obtidos o que corresponde a uma forte probabilidade de codificarem para domínios transmembranares. Finalmente três sequências, B2-2, B3-1 e B3-23, apresentaram resultado positivo em ambas as análises, quer a nível de homologia na base de dados quer a nível da previsão de domínios transmembranares, o que reforça a possibilidade de se tratarem de sequências semelhantes às codificadas pelo gene *env*, característico dos retrovirus. Adicionalmente foi identificado um dos motivos conservados do gene *env* no clone B2-20, LPTL, (Lerat and Capy 1999) numa zona prevista para codificação dum domínio transmembranar pela análise do software TMpred.

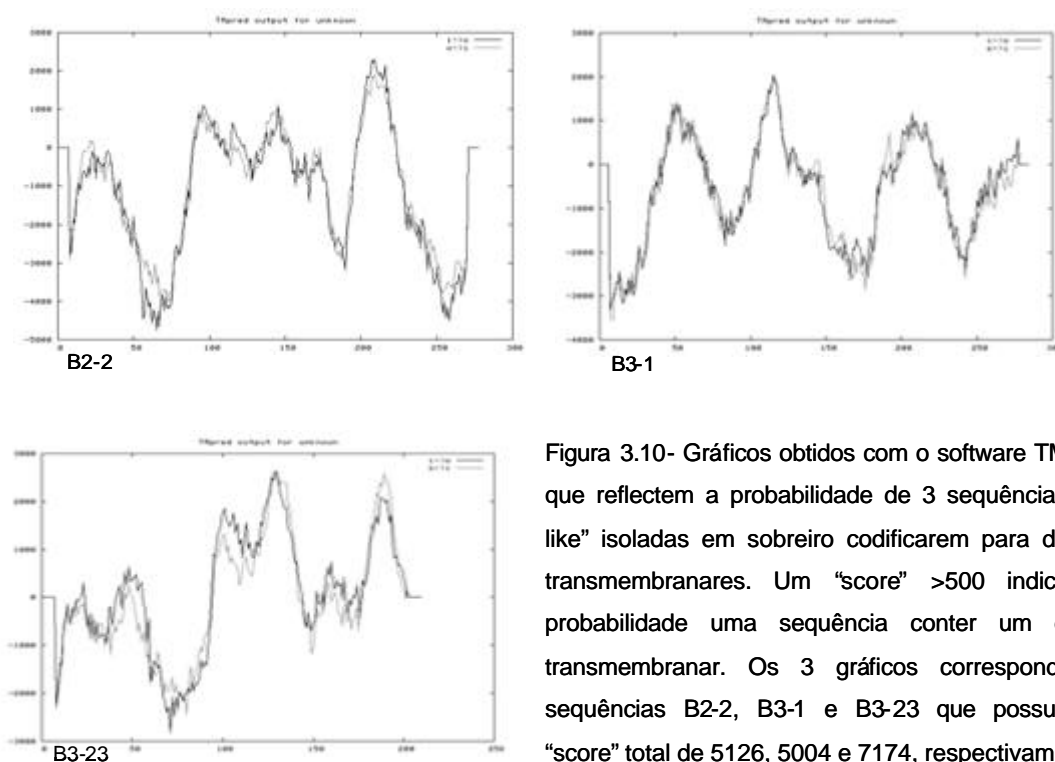


Figura 3.10- Gráficos obtidos com o software TMpred e que reflectem a probabilidade de 3 sequências “*env-like*” isoladas em sobreiro codificarem para domínios transmembranares. Um “score” >500 indica forte probabilidade uma sequência conter um domínio transmembranar. Os 3 gráficos correspondem às sequências B2-2, B3-1 e B3-23 que possuem um “score” total de 5126, 5004 e 7174, respectivamente.

Para determinar possíveis relações entre as sequências isoladas em *Q. suber* e sequências do tipo “*envelope*” de outros organismos, efectuou-se uma análise filogenética com mais algumas sequências do tipo “*envelope*” provenientes de vírus, *HIV-1* e um Iridovirus (*Lymphocystis*), de retrotransposões de plantas: *Bagy-2* (*Ty3-gypsy* de cevada), *SIRE-1* (*Ty1-copia* de soja), *Calypso-1* (*Ty3-gypsy* de soja) e também de *Arabidopsis thaliana*. A análise das sequências de aminoácidos foi realizada pelo método neighbor joining (Saitou and Nei 1987) e o dendograma obtido está representado na figura 3.11.

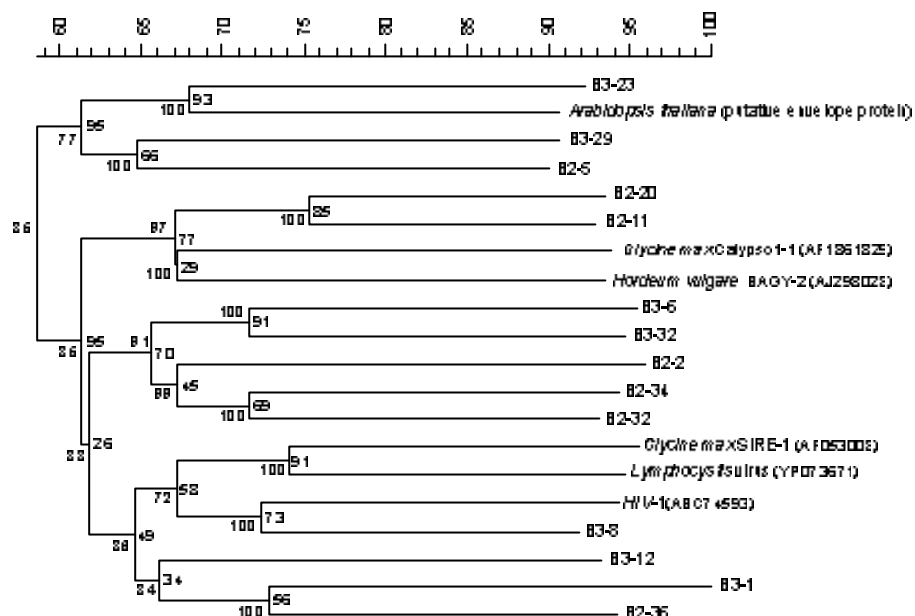


Figura 3.11- Análise filogenética de sequências “env-like” pelo método neighbor joining (Saitou and Nei 1987). O dendograma obtido resulta de um alinhamento das sequências isoladas em *Q.suber* juntamente com sequências previamente identificadas noutras espécies, utilizando o software Bionumerics (versão 3.5). Os 10 clones isolados em *Q. suber* designam-se B2-2, 5, 11, 20, 32, 34, 36 e B3-1, 6, 8, 12, 23, 29 e 32. Os números em cada nó correspondem aos valores percentuais de “bootstrap” relativos a 1000 réplicas e os números em negrito representam as relações cofenéticas, suportando a estabilidade de cada nó. Os números de acesso ao *GenBank* estão associados a cada sequência.

Analisando o dendograma obtido verifica-se que a homologia entre as sequências de sobreiro nunca atinge valores superiores a 76%, valor partilhado pelos clones B2-11 e 20. Adicionalmente destaca-se a existência de dois “clusters” de sequências, ambos suportados por um valor de “bootstrap” de 95%. O primeiro agrupa as sequências B2-5, B3-23 e 29 com *Arabidopsis* e o segundo “cluster” agrupa as restantes sequências usadas nesta análise. Para uma análise mais detalhada o segundo “cluster” pode subdividir-se em três “sub-clusters”. Um deles agrupa apenas sequências isoladas em sobreiro, suportados por um “bootstrap” de 70% e partilhando uma homologia mínima de 66% entre si. Outro “sub-cluster” agrupa dois clones de sobreiro, B2-11 e 20, com dois retroelementos *Ty3-gypsy*, nomeadamente *Calypso-1* de soja (*Glycine max*) e *Bagy-2* de cevada (*Hordeum vulgare*). Estas 4 sequências partilham um mínimo de 67% de homologia a nível dos seus aminoácidos e a relação é suportada por 77% de “bootstrap”. Finalmente o último “sub-cluster” é constituído pelos clones B3-1, 8, 12, e B2-36, clones isolados de 2 virus, *HIV-1* e *Lymphocystis virus*, juntamente com um retroelemento *Ty1-copia* de soja (*Glycine max*), *SIRE-1*.

4. Padrão de distribuição de sequências “*env*-like” no genoma de *Quercus suber* por PCR *in situ*

Com o objectivo de determinar o padrão de distribuição de uma sequência “*env*-like” de sobreiro foi escolhida a sequência B3-1 previamente isolada neste trabalho devido à sua dimensão e homologia detectada com outras sequências “*env*-like” presentes noutros genomas. Desenhou-se um par de “primers” específico para essa sequência e efectuou-se uma amplificação por PCR em DNA genómico, para testar a eficiência dos mesmos. A amplificação decorreu com uma temperatura de “annealing” de 50°C e obteve-se uma banda de DNA com 480bp (Figura 3.12), como esperado.

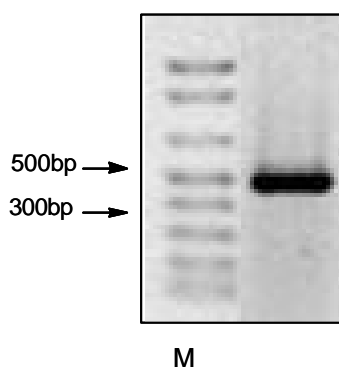


Figura 3.12- Amplificação de uma sequência “*env*-like” em DNA genómico de *Q. suber* pela técnica de PCR. A coluna M corresponde ao marcador 1Kb plus. A amplificação gerou o produto com o tamanho esperado, 480bp, confirmando a especificidade dos “primers” utilizados.

A amplificação da sequência foi efectuada pelo método de C-PRINS em cromossomas e núcleos isolados de ápices radiculares de *Q. suber*. A temperatura usada de “annealing” foi a mesma que na amplificação da sequência de DNA genómico.

Nos núcleos corados com DAPI (figura 3.13-A,C), é possível distinguir o nucléolo com menor intensidade de coloração e blocos perinucleolares de intensa coloração azul que provavelmente correspondem aos *loci* major de rDNA 45S. A sequência usada como sonda apresenta uma marcação fina de aspecto pulverizado e uniformemente distribuída por todo o núcleo, excepto na zona nucleolar onde se detecta menos intensidade, possivelmente devido à menor concentração de cromatina nesta zona (figura 3.13-B,D).

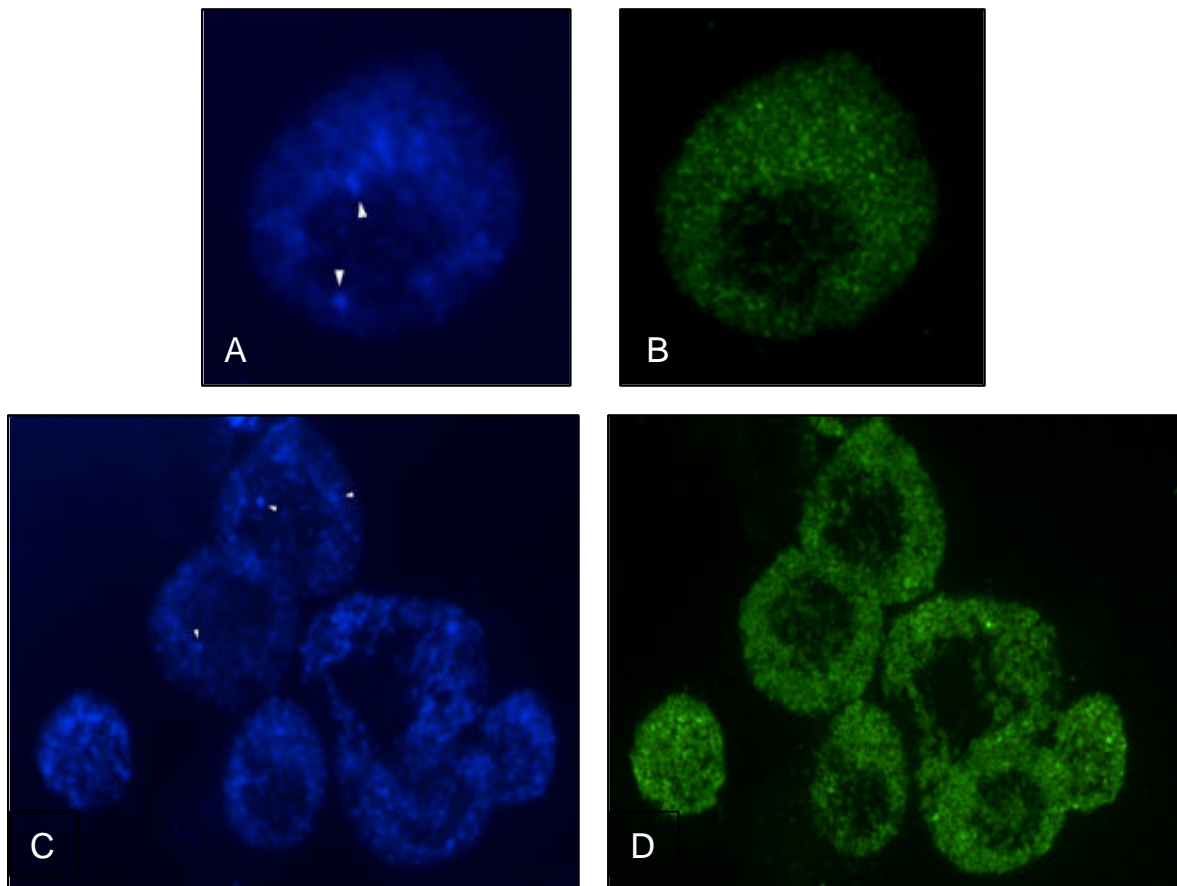


Figura 3.13 - Amplificação de uma sequência “*env*-like” (verde) por C-PRINS em núcleos isolados de ápices radiculares de *Q. suber*. A coloração dos núcleos com DAPI (A,C) permite distinguir o nucléolo, com menor intensidade de coloração, e blocos perinucleolares de intensa coloração azul que provavelmente correspondem aos *loci* major de rDNA 45S, NOR1 (A,C–setas), apesar do estado degradado de alguns núcleos (C). A sequência “*env*” exibe uma marcação fina, de aspecto pulverizado e uniformemente distribuída por todo o núcleo, excepto na zona nucleolar onde se detecta menos intensidade, provavelmente devido à menor concentração de cromatina nesta zona (B,D).

Capítulo 4. Discussão

A espécie *Quercus suber*, é uma das mais importantes espécies de árvores no sudoeste mediterrânico. Faz parte da vegetação natural da península Ibérica, sendo espontânea em muitos locais de Portugal e Espanha. A nível económico é particularmente importante, sendo desta espécie que se extrai a cortiça, da qual Portugal é o maior produtor a nível mundial. A nível ecológico tem um papel preponderante na nossa flora, sendo mais abundante no centro e especialmente sul do país, constituindo os montados. Estes oferecem uma boa protecção do solo e pela sua fraca cobertura sub-arbustiva são um forte aliado na luta contra os incêndios.

Na área da genética molecular, os elementos transponíveis são uma área emergente começando a perceber-se a sua extrema importância no genoma eucariota. No presente trabalho inicia-se o estudo de retrotransposições LTR no genoma de *Quercus suber*, com uma primeira identificação de vários destes elementos, sua análise e distribuição no genoma.

1. Retrotransposições LTR em plantas: transmissão vertical, horizontal ou ambas?

- Análise de RVT's *Ty1-copia*

Os retrotransposições LTR *Ty1-copia* são ubíquos nos organismos eucariotas, incluindo fungos, animais e plantas (Flavell et al. 1992; Voytas et al. 1992). A enzima transcriptase reversa (RVT), responsável pela síntese de novas moléculas de cDNA do retrotransposão, possui um elevado grau de conservação a nível da sua sequência (Doolittle et al. 1989) sendo muitas vezes usada como sonda em estudos filogenéticos e no mapeamento físico de retrotransposições em genomas. Neste estudo foram amplificados com sucesso os domínios de transcriptases reversas de retrotransposições *Ty1-copia* no genoma de sobreiro, *Quercus suber*. A estratégia seguida envolveu o uso de “primers” degenerados desenhados para motivos conservados neste gene, nomeadamente os motivos QMDVKT e YVDDM. Isolou-se uma população de 10 clones que apresentava alguma variedade entre si, revelando uma homologia de 63 a 93% nas suas sequências, ao nível dos aminoácidos. Este facto corrobora o que tem sido descrito por alguns autores em plantas, incluindo briófitos, pteridófitos, gimnospermas e angiospermas (Favell et al. 1992), contrastando com a relativamente baixa heterogeneidade detectada em *Saccharomyces cerevisiae* e *Drosophila melanogaster*. Não foi possível definir famílias de retrotransposições LTR em *Quercus suber*, uma vez que para tal é necessário que as RVT's isoladas apresentem uma homologia a nível das sequências de aminoácidos igual ou superior a 90% (Bowen and McDonald 2001) e apenas duas sequências (C1-6 e 17) cumpriram esse requisito, partilhando 93% de homologia entre si. Uma hipótese para esta variabilidade é o próprio mecanismo pelo qual

os retrotransposões originam novas cópias. A síntese das novas moléculas de cDNA são produzidas pela RVT utilizando os transcritos como molde. A RVT é uma enzima que não possui actividade exonucleotídica no sentido 3'-5' (ref) e como tal não corrige mutações, no caso de as originar aquando da síntese de novas moléculas, sendo responsável por parte da variabilidade que se observa nas sequências isoladas. Relativamente à análise filogenética com sequências de RVT's isoladas noutras espécies vegetais, verifica-se que as sequências estão agrupadas em dois grandes "clusters", dos quais não faz parte o clone isolado em *Diospyros kaki*, que constitui um "outgroup", partilhando apenas 56% de homologia com as restantes. No primeiro grande cluster, suportado por um valor de bootstrap de 79%, existem sequências provenientes de duas divisões taxonómicas distintas: Cycadophyta (*Cycas revoluta*) e Magnoliophyta (restantes clones). Adicionalmente, verifica-se a existência de "sub-clusters" cuja homologia entre as sequências de RVT's reflecte a proximidade taxonómica das espécies nas quais foram isoladas. Trigo (*Triticum aestivum*) e centeio (*Secale cereale*), da família Poaceae partilham 69% de homologia e definem um "sub-cluster"; batata (*Solanum tuberosum*) e tomate (*Lycopersicum esculentum*) da família Solanaceae e uma espécie próxima, melão (*Cucumis melo*), da família Cucurbitaceae partilham entre 70-73% de homologia entre si e definem outro "sub-cluster"; 3 clones de sobreiro, C1-6, 17 e C2-33 partilham entre 74-84% de homologia e definem mais um "sub-cluster". Estes dados estão de acordo com a teoria que postula que os retrotransposões LTR constituem inserções ancestrais cuja dispersão é feita por transmissão vertical. Este tipo de transmissão associado ao mecanismo de transposição característico destes elementos e a eventuais perdas que possam ocorrer nos genomas, gera um padrão em que a homologia detectada entre clones ou famílias de clones isolados na mesma espécie são mais semelhantes entre si que entre espécies, particularmente se estas forem evolutivamente distantes (Konieczny et al. 1991). No entanto, a análise do último "sub-cluster" que agrupa clones de sobreiro, C1-2 e 32, leguminosas (*G. max* e *P. sativum*), *I. batatas* da família Convolvulaceae e uma espécie ancestral na evolução das plantas, *Cycas revoluta*, da divisão Cycadophyta conduz a outra perspectiva. Este conjunto partilha uma homologia entre 69 e 81% e é suportado por um valor de "bootstrap" de cerca de 70%. Adicionalmente, um clone de sobreiro, C1-19, possui maior homologia com clones isolados em laranjeira (*Citrus sinensis*) e em espargos (*Asparagus officinalis*), das famílias Rutaceae e Asparagaceae, respectivamente, em relação a 4 clones de sobreiro, C1-3, 7, 20 e C2-5, englobados num "cluster" suportado por 96% de bootstrap. Estes resultados mostram elevada homologia entre sequências evolutivamente distantes. Por exemplo, 69% de homologia entre uma sequência isolada em *Cycas revoluta*, separada de *Quercus suber* por 360 milhões de anos, altura em que gimnospérmicas e angiospérmicas divergiram. Os retrotransposões incorporam mutações a uma taxa um milhão de vezes superior à referida

para genes nucleares como consequência do seu mecanismo de transposição, gerando níveis elevados de variabilidade entre sequências (Gojobori et al. 1990). Como tal é altamente improvável que este grau de homologia fosse conservado ao longo de tantos anos de evolução unicamente por transmissão vertical. Assim, parece evidente que fenómenos de transmissão horizontal tenham desempenhado um papel importante na evolução e dispersão dos retrotransposões entre organismos distintos, especialmente no reino vegetal. De facto estudos recentes apontam para a conjugação de ambos os tipos de transmissão na dispersão deste tipo de elementos, ao invés de optarem por uma única força condutora (Flavell et al. 1992; Pearce 2007; Yano et al. 2005). A transmissão horizontal pode ser feita por espécies que desempenham o papel de vector como, por exemplo, insectos ou aves, ou através de polinização cruzada. Num estudo recente foram inclusivamente descritas trocas citoplasmáticas entre duas espécies do género *Quercus*, mais precisamente *Quercus suber* e *Quercus ilex* (Belahbib et al. 2001).

- Análise de sequências “env-like” em retrotransposões *Ty1-copia* de *Q. suber*

O gene envelope, *env*, é característico dos virus e está associado à sua capacidade infecciosa. Codifica para glicoproteínas transmembranares que são integradas no invólucro lipídico das partículas virais e reconhecidas por receptores celulares, facilitando a fusão entre a membrana do virião e da célula, permitindo a entrada do genoma viral na mesma (Temin 1980). Porém tem sido identificado em alguns retrotransposões LTR de plantas, *Ty3-gypsy* ou *Ty1-copia*, nomeadamente na divisão Magnoliophyta (Angiospérmicas). São exemplos do tipo *Ty3-gypsy* a família *Athila* de *Arabidopsis thaliana* (Wright and Voytas 1998a) com os homólogos *Cyclops* (Chavanne et al. 1998) e *Calypso* (Peterson-Burch et al. 2000) em ervilha e soja, e o elemento *SIRE-1* de soja, *Ty1-copia* (Laten et al. 2003; Laten et al. 1998), com uma família de homólogos em *Arabidopsis thaliana* designada *Endovir* (Peterson-Burch et al. 2000). Recentemente foi também identificada a presença deste gene no genoma de uma gimnospérmica (divisão Spermatophyta), *Pinus pinaster* (Rocheta comunicação pessoal). Quando presente em retrotransposões LTR, o gene *env* tem a mesma localização que nos retrovirus, a jusante do gene *pol* e a montante da LTR 3' (Laten et al. 2003). Neste trabalho foram isolados com sucesso clones parciais deste gene em retrotransposões *Ty1-copia* no genoma de *Quercus suber*. Os clones resultaram de uma amplificação de DNA por PCR entre um domínio conservado da *RnaseH* (RTKHID/E) e a LTR 3', seguindo a estratégia de Pearce (Pearce et al. 1999). De um total de 17 clones isolados, 14 (82%) apresentaram forte probabilidade de codificarem para sequências do tipo “env-like”. Destas catorze, cinco apresentam homologia a nível dos aminoácidos com sequências do tipo “env” de outros organismos, nomeadamente *HIV-1* e *SRV-2* (Simian

retrovirus), detectada por pesquisa na base de dados *GenBank*. Utilizando o software TMpred, que prevê a existência de domínios transmembranares e a sua orientação na membrana celular (Hofmann and Stoffel 1993) verificou-se que doze das catorze sequências exibiram valores superiores a 500, o que traduz uma elevada probabilidade de codificarem para domínios transmembranares. Finalmente, três sequências apresentaram resultado positivo em ambas as análises, quer a nível de homologia na base de dados quer a nível da previsão de domínios transmembranares, o que reforça a possibilidade de se tratarem de efectivamente de sequências *env*. Dentro da população de clones isolados em sobreiro verifica-se que a homologia entre as sequências, ao nível dos aminoácidos, atinge um valor máximo de 76% entre os clones B2-11 e 20. Este é um valor inesperado e é considerado elevado para este tipo de sequências que têm sido descritas como pouco conservadas, mesmo em retrovírus (Chavanne et al. 1998; Peterson-Burch et al. 2000). Até à data não se sabe qual a vantagem que a presença destas sequências confere ao genoma da planta, contudo uma maior homologia poderá indicar algum tipo de preservação importante para o hospedeiro cuja função ainda é desconhecida. Adicionalmente, o clone B2-20 apresenta o motivo LTPL que está presente na maioria dos *env* de retrovirus e retrotransposições e que se supõe ter um papel importante na adsorção e penetração do virus na célula hospedeira (Lerat and Capy 1999).

Com o objectivo de determinar possíveis relações entre sequências *env* de vários organismos, fez-se uma análise filogenética incluindo as sequências isoladas neste trabalho juntamente com sequências de virus e de retrotransposições LTR de plantas (Figura 3.10). Destaca-se a existência de dois “clusters” de sequências, ambos suportados por um valor de “bootstrap” de 95%: o primeiro agrupa os clones de sobreiro B2-5, B3-23 e 29 com um clone isolado em *Arabidopsis* e o segundo “cluster” agrupa os restantes clones usados nesta análise. Este pode ainda ser subdividido em três “sub-clusters”, para uma análise mais detalhada. Um deles agrupa apenas clones isolados em sobreiro, suportados por um “bootstrap” de 70% e partilhando uma homologia mínima de 66% entre si. Outro “sub-cluster” agrupa dois clones de sobreiro, B2-11 e 20, com dois retroelementos *Ty3-gypsy*, nomeadamente *Calypso-1* de soja (*Glycine max*) e *Bagy-2* de cevada (*Hordeum vulgare*). Estas 4 sequências partilham um mínimo de 67% de homologia a nível dos seus aminoácidos e a relação é suportada por 77% de “bootstrap”. Finalmente o último “sub-cluster” é constituído pelos clones B3-1, 8, 12, e B2-36, clones isolados de 2 virus, HIV-1 e Lymphocystis virus, juntamente com um retroelemento *Ty1-copia* de soja (*Glycine max*), *SIRE-1*. Desta análise podemos concluir que a homologia entre clones isolados em sobreiro, por exemplo no “sub-cluster” constituído pelos clones B2, 32, 34 e B3-6,32 relativamente a clones isolados em virus ou noutros retrotransposições de plantas, suporta a hipótese de que a transmissão destes elementos é essencialmente feita por via vertical. A variabilidade entre

as sequências isoladas pode ser explicada como consequência do seu mecanismo de transposição (Konieczny et al. 1991) além da pouca conservação que tem sido descrita para o gene, referida anteriormente. Por outro lado, é-nos sugerida a presença de uma sequência *env* ancestral à separação dos retrotransposições LTR em *Ty1-copia* e *Ty3-gypsy*, pela homologia detectada entre os clones B2-11 e 20, do tipo *Ty1-copia* e dois retroelementos *Ty3-gypsy*, nomeadamente *Calypso-1* de soja e *Bagy-2* de cevada. Adicionalmente temos o “sub-cluster” formado por clones de sobreiro, de vírus (HIV-1 e Lymphocystis virus) e de um retrotransposição LTR *Ty1-copia* de soja, *SIRE-1*. Recentemente foi descrito um estudo que afirma que o elemento *SIRE-1* adquiriu o gene *env* de forma independente, pela sua proximidade com uma família de elementos *Ty1-copia* que não o possui (Pearce 2007). Nesta análise filogenética, na qual se incorporou um clone de *SIRE-1*, este elemento partilha 74% de homologia com um clone de um vírus, sugerindo que a aquisição do gene *env* poderá ter sido feita pela incorporação do gene de um vírus que foi transmitido horizontalmente. O agrupamento de outro clone de vírus, HIV-1, neste “sub-cluster”, juntamente com um clone de sobreiro pode indicar que existe uma linhagem homóloga ao *SIRE-1* no genoma de sobreiro. O gene *env* está presente em ambas as famílias *Ty3-gypsy* e *Ty1-copia*, como foi descrito acima, o que sugere que se foi transmitido por via horizontal, a aquisição deste pelos retrotransposições terá que ter ocorrido no mínimo duas vezes e duma forma independente (Kumar 1998). A detecção de homologia entre sequências *env* de plantas e retrovírus de organismos taxonomicamente distantes tem sido descrita por outros autores, nomeadamente em *Gossypium* em relação ao vírus SIV (Ghany and Zaki 2002). Além disso, a existência de diferentes “clusters” que contêm clones de sobreiro pode também significar que o gene *env* evoluiu de forma independente, dentro da mesma planta, mesmo que seja transferido verticalmente. À excepção do clone B2-20, referido anteriormente, não foram identificados os motivos conservados dos *env* de retrovírus e retrotransposições, KRG e LTPL, em mais nenhum dos clones isolados. Se estes motivos existiram no gene *env* de retrotransposições no passado e de alguma forma os perderam, significa que estão a ser sujeitos a algum tipo de pressão selectiva negativa. Assim, embora a parede celular das plantas seja um constrangimento físico à proliferação deste tipo de elementos, parece não ser o único obstáculo a contornar em plantas (Kumar and Bennetzen 1999). Noutra perspectiva, na ausência de uma função conhecida nas plantas, o gene *env* poderá ter acumulado mutações ao longo do tempo, tornando-se num pseudogene sem função aparente (Kumar and Bennetzen 1999).

2. Contribuição das sequências de RVT's e “env-like” para a organização do genoma de *Q. suber*

Os retrotransposões LTR são elementos genéticos móveis ubíquos no reino vegetal e estão presentes num elevado número de cópias, constituindo uma fracção significativa do genoma de inúmeras espécies. Em milho, 50-80% do genoma nuclear é atribuído a estes elementos (SanMiguel and Bennetzen 1998), em arroz, 22% (Ma et al. 2004) e mais de 50% no genoma de cevada, onde uma única família de retrotransposões, a família *BARE-1* representa por si só, 5% do seu genoma (Vicient et al. 1999b). Aliando-se a este facto, muitos estudos têm sido conduzidos acerca da maneira como estes elementos se distribuem e organizam no genoma de plantas. A sequência da RVT é frequentemente utilizada como sonda uma vez que apresenta um elevado grau de conservação. Como tal neste trabalho foram estudadas duas sequências de RVT's no genoma de *Quercus suber* por PCR *in situ*, uma *gypsy-like* e uma *copia-like* de 650 e 250bp, respectivamente, assim como uma sequência “env-like” de 480bp. O padrão de marcação obtido para as sequências de RVT's de ambas as famílias diferiu. A RVT do tipo *Ty1-copia* exibiu uma marcação uniforme por todo o genoma, não sendo detectada nenhuma região de marcação mais intensa. Este resultado tem sido descrito em pteridófitos, gimnospermas e angiospermas (Brandes et al. 1997b). O mesmo padrão é encontrado também em *Pinus elliotii* (Kamm et al. 1996), aveia (Katsiotis et al. 1996), cevada (Waugh et al. 1997), centeio e fava. Em cebola os retrotransposões *Ty1-copia* estão também presentes por todo o genoma, em zonas eucromáticas, mas encontram-se concentrados sobretudo em regiões heterocromáticas terminais. No genoma de milho estes são identificados em zonas intergénicas, num arranjo “nested” (inseridos uns dentro dos outros) (SanMiguel et al. 1996b), enquanto que em sorgo, estão presentes em regiões eucromáticas (Miller et al. 1998). Relativamente à sequência do tipo *Ty3-gypsy* obteve-se uma marcação dispersa pelo genoma, com um aspecto algo “pontado” e presente em todos os cromossomas. Contudo observou-se também um sinal de maior intensidade numa zona heterocromática, o NOR1. Este é o major *locus* de rDNA 45S em *Quercus suber*, que ocupa uma posição sub-terminal, em oposição ao *locus* minor com uma posição pericentromérica (NOR2) noutro par de cromossomas (Zoldos et al. 1999). A hibridação *in situ* sequencial com a sonda pTa71, específica para genes de rRNA 45S, veio confirmar a co-localização da RVT com o NOR major. No NOR2, que ocupa uma posição pericentromérica, não foi possível detectar a co-localização destas duas sequências sendo a distribuição da RVT semelhante à encontrada no resto do genoma. Num par de cromossomas observa-se uma marcação mais intensa na zona centromérica, também caracteristicamente heterocromática e rica em sequências repetitivas. Em várias espécies

vegetais tem sido descrita a presença de retroelementos em regiões centroméricas e pericentroméricas. Em cevada o retroelemento *Ty3-gypsy* Cerebe, encontra-se nestes domínios e em sorgo elementos *Ty3-gypsy* são exclusivos dos centrómeros (Miller et al. 1998). Também em *Arabidopsis* a família *Athila* localiza-se apenas nos cromocentros (Lin et al. 1999).

Relativamente à sonda *env*, obteve-se uma marcação semelhante à sonda do tipo *Ty1-copia*, dispersa pelos núcleos, com um aspecto ponteadado. À semelhança do que sucedeu com as preparações citológicas onde se usou a sonda *Ty1-copia*, não se obtiveram cromossomas, pelo que não é possível inferir o resultado em zonas teloméricas ou centroméricas.

O padrão de distribuição disperso e muito pulverizado que se obtém para ambos os tipos de retroelementos no genoma de sobreiro parece indicar que não há zonas preferenciais de inserção destes elementos, consequência provável do seu mecanismo de transposição. Este é feito pela transcrição da sua molécula de DNA, originando um transcrito que é convertido em cDNA pelas enzimas RVT e RnaseH, que pode ser integrado num local diferente do genoma. No entanto, este mecanismo pode ser potencialmente invasivo e causar efeitos deletérios no genoma hospedeiro e, consequentemente, no próprio retrotransposição. Como tal parece que estes elementos adoptaram estratégias que minimizam os efeitos deletérios causados pela sua inserção como é o caso de se inserirem em zonas heterocromáticas, pouco ricas em genes e geralmente associadas a um estado de inactivação transcricional.

Por outro lado, o padrão disperso observado, com marcações muito finas parece também ser o reflexo da ausência de “clusters” destes elementos no genoma, para além da zona dos NOR's major, onde se detecta uma acumulação deste retroelemento. Uma vez que em sobreiro estes genes são funcionais, o que se pode comprovar pela sua associação ao nucléolo, é de supor que este retroelemento se encontre inserido nas zonas intergénicas presentes neste *locus*, ou, em zonas génicas de apenas algumas cópias. Os genes de rDNA 45S existem em milhares de cópias por genoma encontrando-se estas separadas por sequências que não são transcritas (ETS). Se algumas destas cópias forem inactivadas pela presença destes elementos, o *locus* continuará a ter cópias funcionais em número suficiente para assegurar o funcionamento celular. De facto, em várias espécies do filo Cnidaria e em artrópodes foi descrita uma família de retrotransposições, que se insere em zonas mais conservadas da sequência de rDNA 45S (Kojima and Fujiwara 2005; Kojima et al. 2006).

3. Conclusões finais

Os resultados obtidos mostram que a espécie *Quercus suber* possui retrotransposições LTR *copia* e *gypsy-like* dispersos pelo seu genoma, ausentes de regiões heterocromáticas, à exceção do elemento *gypsy-like* que exibe uma marcação mais intensa no *locus* de rDNA 45S NOR1 e no centrómero de um par de cromossomas. Esta distribuição pode constituir uma estratégia de adaptação dos retrotransposições LTR para minimizar os efeitos deletérios no genoma hospedeiro, assegurando a sua manutenção e propagação no mesmo.

As análises filogenéticas das sequências isoladas neste trabalho juntamente com sequências isoladas noutras espécies permite concluir que embora a transmissão vertical seja a força condutora da dispersão dos retrotransposições LTR, durante a história evolutiva destes elementos terá ocorrido transmissão horizontal.

Qualquer dos cenários apresentados sugere que os retrotransposições poderão ter um papel mais importante do que se julgava inicialmente. Estudos futuros nesta área irão certamente fornecer provas do mesmo.

Referências bibliográficas

- Belahbib, N., M.H. Pemonge, A. Ouassou, H. Sbay, A. Kremer, and R.J. Petit. (2001). Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. *Mol Ecol* **10**: 2003-2012.
- Bennett, M.D. and I.J. Leitch. (2005). Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Ann Bot (Lond)* **95**: 45-90.
- Bennetzen, J.F. (2000). Transposoable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **42**: 251-269.
- Bennetzen, J.L. (1996). The contributions of retroelements to plant genome organization, function and evolution. *Trends Microbiol* **4**: 347-353.
- Bennetzen, J.L. (2002). Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**: 29-36.
- Boeke, J.D. (1989). Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.* **43**: 403-434.
- Bowen, N.J. and J.F. McDonald. (2001). Drosophila euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res* **11**: 1527-1540.
- Brandes, A., J.S. Heslop-Harrison, A. Kamm, S. Kubis, R.L. Doudrick, and T. Schmidt. (1997a). Comparative analysis of the chromosomal and genomic organization of Ty1-copia-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol Biol* **33**: 11-21.
- Brandes, A., J.S. Heslop-Harrison, A. Kamm, S. Kubis, R.L. Doudrick, and T. Schmidt. (1997b). Comparative analysis of the chromosomal and genomic organization of Ty1-copia-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol. Biol.* **33**: 11-21.
- Carrington, J.C., K.D. Kasschau, S.K. Mahajan, and M.C. Schaad. (1996). Cell-to-Cell and Long-Distance Transport of Viruses in Plants. *Plant Cell* **8**: 1669-1681.
- Chavanne, F., D.-X. Zhang, M.-F. Liaud, and R. Cerff. (1998). Structure and evolution of *Cyclops*: a novel giant retrotransposon of the Ty3/Gypsy family highly amplified in pea and other legume species. *Plant Mol. Biol.* **37**: 363-375.
- Cinti, C., S. Santi, and N.M. Maraldi. (1993). Localization of single copy gene by PRINS technique. *Nucleic Acids Res* **21**: 5799-5800.
- Doolittle, R.F., D.-F. Feng, M.S. Jonhson, and M.A. McClure. (1989). Origins and evolutionary relationships of retroviruses. *Q. Rev. Biol.* **64**: 1-30.
- Doolittle, R.F. and D.F. Feng. (1992). Tracing the origin of retroviruses. *Curr Top Microbiol Immunol* **176**: 195-211.
- Favell, A.J., E. Dunbar, R. Anderson, S.R. Pearce, and R. Hartley. (1992). Ty-1 copia group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Res* **20**: 3639-3644.
- Flavell, A.J., E. Dunbar, R. Anderson, S.R. Pearce, R. Hartley, and A. Kumar. (1992). Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Res* **20**: 3639-3644.
- Ghany, A. and E. Zaki. (2002). Cloning and sequencing of an *envelope*-like gene in *Gossypium*. *Planta* **216**: 351-353.

- Gojobori, T., E.N. Moriyama, and M. Kimura. (1990). Molecular clock of viral evolution, and the neutral theory. *Proc Natl Acad Sci U S A* **87**: 10015-10018.
- Grandbastien, M.-A. (1998). Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* **3**: 181-187.
- Hawkins, J.A., H. Kim, J.D. Nason, and R.A. Wing. (2006). Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* **16**: 1252-1261.
- Hofmann, K. and W. Stoffel. (1993). TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**: 166.
- Hull, R. (2001). Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. *Arch. Virol.* **146**: 2255-2261.
- Jiang, N., Z. Bao, S. Temnykh, Z. Cheng, J. Jiang, R.A. Wing, S.R. McCouch, and S.R. Wessler. (2002). *Dasheng*: A Recently Amplified Nonautonomous Long Terminal Repeat Element That Is a Major Component of Pericentromeric Regions in Rice. *Genetics* **161**: 1293–1305.
- Kadandale, J.S., Y. Tunca, and A.T. Tharapel. (2000). Chromosomal localization of single copy genes SRY and SOX3 by primed in situ labeling (PRINS). *Microb Comp Genomics* **5**: 71-74.
- Kamm, A., R.L. Doudrick, J.S. Heslop-Harrison, and T. Schmidt. (1996). The genomic and physical organization of *Ty1-copia*-like sequences as a component of large genomes in *Pinus elliotii* var. *elliotii* and other gymnosperms. *Proc. Natl. Acad. Sci.* **93**: 2708–2713.
- Kapuscinski, J. (1995). DAPI: a DNA-specific fluorescent probe. *Biotech Histochem.* **70(5)**: 220-223
- Katsiotis, A., T. Schmidt, and J.S. Heslop-Harrison. (1996). Chromosomal and genomic organization of Ty1-copia-like retrotransposon sequences in the genus *Avena*. *Genome* **39**: 410-417.
- Kojima, K.K. and H. Fujiwara. (2005). Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol Biol Evol* **22**: 2157-2165.
- Kojima, K.K., K. Kuma, H. Toh, and H. Fujiwara. (2006). Identification of rDNA-specific non-LTR retrotransposons in Cnidaria. *Mol Biol Evol* **23**: 1984-1993.
- Konieczny, A., D.F. Voytas, M.P. Cummingst, and F.M. Ausubel. (1991). A Superfamily of *Arabidopsis thaliana* Retrotransposons. *Genetics* **127**: 801-809.
- Kumar, A. (1998). The evolution of plant retroviruses: moving to green pastures. *Trends Plant Sci.* **3**: 371-374.
- Kumar, A. and J.L. Bennetzen. (1999). Plant Retrotransposons. *Annu. Rev. Genet.* **33**: 479-532.
- Kunze, R., H. Saedler, and W.E. Lonig. (1997). Plant transposable elements. *Adv Bot Res*: 331-470.
- Laten, H.M. (1999). Phylogenetic evidence for *Ty1-copia*-like endogenous retroviruses in plant genomes. *Genetica* **107**: 87–93.
- Laten, H.M., E.R. Havecker, L.M. Farmer, and D.F. Voytas. (2003). SIRE1, an Endogenous Retrovirus Family from Glycine max, Is Highly Homogeneous and Evolutionarily Young. *Mol. Biol. Evol.* **20**: 1222–1230.
- Laten, H.M., A. Majumdar, and E.A. Gaucher. (1998). *SIRE-1*, a *copia Ty1*-like retroelement from soybean, encodes a retroviral envelope-like protein. *Proc. Natl. Acad. Sci.* **95**: 6897–6902.

- Lerat, E. and P. Capy. (1999). Retrotransposons and Retroviruses: Analysis of the Envelope Gene. *Mol. Biol. Evol.* **16**: 1198–1207.
- Lin, X., S. Kaul, S. Rounsley, T.P. Shea, M.I. Benito, C.D. Town, C.Y. Fujii, T. Mason, C.L. Bowman, M. Barnstead, T.V. Feldblyum, C.R. Buell, K.A. Ketchum, J. Lee, C.M. Ronning, H.L. Koo, K.S. Moffat, L.A. Cronin, M. Shen, G. Pai, S. Van Aken, L. Umayam, L.J. Tallon, J.E. Gill, M.D. Adams, A.J. Carrera, T.H. Creasy, H.M. Goodman, C.R. Somerville, G.P. Copenhaver, D. Preuss, W.C. Nierman, O. White, J.A. Eisen, S.L. Salzberg, C.M. Fraser, and J.C. Venter. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**: 761-768.
- Ma, J., K.M. Devos, and J.L. Bennetzen. (2004). Analyses of LTR-Retrotransposon Structures Reveal Recent and Rapid Genomic DNA Loss in Rice. *Genome Res.* **14**: 860-869.
- Malik, H.S., S. Henikoff, and T.H. Eickbush. (2000). Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* **10**: 1307-1318.
- Matsuoka, Y. and K. Tsunewaki. (1999). Evolutionary dynamics of Ty1-copia group retrotransposons in grass shown by reverse transcriptase domain analysis. *Mol Biol Evol* **16**: 208-217.
- Miller, J.T., F. Dong, S.A. Jackson, J. Song, and J. Jiang. (1998). Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* **150**: 1615-1623.
- Mumford, R.A., I. Barker, and K.R. Wood. (1996). The biology of tospoviruses. *Ann. Appl. Biol.* **128**: 159-183.
- Neumann, P., D. Pozarkova, A. Koblizkova, and J. Macas. (2005). PIGY, a new plant envelope-class LTR retrotransposon. *Mol Genet Genomics* **273**: 43-53.
- Pearce, S.R. (2007). SIRE-1, a putative plant retrovirus is closely related to a legume TY1-copia retrotransposon family. *Cell Mol Biol Lett* **12**: 120-126.
- Pearce, S.R., G. Harrison, D. Li, J. Heslop-Harrison, A. Kumar, and A.J. Flavell. (1996a). The Ty1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. *Mol Gen Genet* **250**: 305-315.
- Pearce, S.R., U. Pich, G. Harrison, A.J. Flavell, J.S. Heslop-Harrison, I. Schubert, and A. Kumar. (1996b). The Ty1-copia group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. *Chromosome Res* **4**: 357-364.
- Pearce, S.R., C. Stuart-Rogers, M.R. Knox, A. Kumar, T.H.N. Ellis, and A.J. Flavell. (1999). Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. *Plant J.* **19**: 711-717.
- Peterson-Burch, B.D., D.A. Wright, H.M. Laten, and D.F. Voytas. (2000). Retroviruses in plants? *Trends Genet.* **16**: 151-152.
- Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Saniyal, H. Kim, K. Collura, D. S. Brar, S. Jackson, R. A. Wing, and O. Panaud. (2006). Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* **16**: 1262-1269.
- Price, Z., F. Dumortier, D.W. MacDonald, and S. Mayes. (2002). Characterisation of copia-like retrotransposons in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* **104**: 860-867.
- Saitou, N. and M. Nei. (1987). The Neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406-425.

- Sambrook, J., E.F. Fritsch, and T. Maniatis. (1989). Molecular cloning: A laboratory manual, 2nd edn. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York*.
- SanMiguel, P. and J.L. Bennetzen. (1998). Evidence that a Recent Increase in Maize Genome Size was Caused by the Massive Amplification of Intergene Retrotransposons. *Annals of Botany* **82 (SupplementA)**: 37-44.
- SanMiguel, P., B.S. Gaut, A. Tikhonov, Y. Nakajima, and J.L. Bennetzen. (1998). The paleontology of intergene retrotransposons of maize. *Nat Genet* **20**: 43–45.
- SanMiguel, P., A. Tikhonov, Y.-K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova, and J.L. Bennetzen. (1996a). Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765-768.
- SanMiguel, P., A. Tikhonov, Y.K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova, and J.L. Bennetzen. (1996b). Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765-768.
- Schwarzacher, T. and P. Heslop-Harrison. 2000. *Practical in situ Hybridization*.
- Selinger, D.A. and V.L. Chandler. (2001). B-Bolivia, an Allele of the Maize b1 Gene with Variable Expression, Contains a High Copy Retrotransposon-Related Sequence Immediately Upstream1. *Plant Physiol.* **125**: 1363-1379.
- Sharma, A.K. and A. Sharma. 1999. *Plant chromossomes: analysis, manipulation and engineering*. . Harwood Academic Publishers, Amsterdam.
- Shirasu, K., A.H. Schulman, T. Lahaye, and P. Shulze-Lefert. (2000). A Contiguous 66-kb Barley DNA Sequence Provides Evidence for Reversible Genome Expansion. *Genome Res.* **10**: 908-915.
- Smith, M.W., D.F. Feng, and R.F. Doolittle. (1992). Evolution by acquisition: the case for horizontal gene transfers. *Trends Biochem Sci* **17**: 489-493.
- Song, S.U., T. Gerasimova, M. Kurkulos, J.D. Boeke, and V.G. Corces. (1994). An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev* **8**: 2046-2057.
- Temin, H.M. (1980). Origin of retroviruses from cellularmovable genetic elements. *Cell* **21**: 599–600.
- van-Regenmortel, M.H.V., C.V. Fauquet, D.H.L. Bishop, E.B. Carsten, and M.K. Estes. 2000. Virus taxonomy: Seventh report of the international committee on taxonomy of viruses. Academic Press, San Diego.
- Vandenheuvel, J.F.J.M., A.W.E. Franz, and F. Vandermlk. 2002. Molecular basis of virus transmission. In *Molecular biology of plant viruses* (ed. C.L. Mandahar), pp. 183-210. Kluwer, Boston.
- Varmus, H. and P. Brown. 1989. Retroviruses. In *Mobile DNA* (eds. D.E. Berg and M.M. Howe), pp. 53-108. American Society for Microbiology, Washington, D.C.
- Vernhettes, S., M.-A. Grandbastien, and J.M. Casacuberta. (1998). The Evolutionary analysis of the *Tnt1* retrotransposon in *nicotiana* species reveals the high variability of its regulatory sequences. *Mol. Biol. Evol.* **15**: 827–836.
- Vicient, C.M., R. Kalendar, K. Anamthawat-Jonsson, and A.H. Schulman. (1999a). Structure, functionality, and evolution of the BARE -1 retrotransposon of barley. *Genetica* **107**: 53-63.
- Vicient, C.M., R. Kalendar, and A.H. Schulman. (2001). Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. *Genome Res* **11**: 2041-2049.

- Vicient, C.M., A. Suoniemi, K. Anamthawat-Jonsson, J. Tanskanen, A. Beharav, E. Nevo, and A.H. Schulman. (1999b). Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus *Hordeum*. *Plant Cell* **11**: 1769-1784.
- Voytas, D.F., M.P. Cummings, A. Konieczny, F.M. Ausubel, and S.R. Rodermel. (1992). Copia-like retrotransposons are ubiquitous among plants. *Proc. Natl. Acad. Sci.* **89**: 7124-7128,.
- Wachtel, S.S. and A.T. Tharapel. (2006). PRINS for the detection of unique sequences. *Methods Mol Biol* **334**: 33-40.
- Waugh, R., K. McLean, A.J. Flavell, S.R. Pearce, A. Kumar, B.B.T. Thomas, and W. Powell. (1997). Genetic distribution of *Bare-1*-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol. Gen. Genet.* **253**: 687-694.
- Wright, D.A., N. Ke, J. Smalle, B.M. Hauge, H.M. Goodmant, and D.F. Voytas. (1996). Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. *Genetics* **142**: 569-578.
- Wright, D.A. and D.F. Voytas. (1998a). Potential retroviruses in plants: *Tat1* Is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. *Genetics* **149**: 703-715.
- Wright, D.A. and D.F. Voytas. (1998b). Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. *Genetics* **149**: 703-715.
- Wright, D.A. and D.F. Voytas. (2002). *Athila4* of *Arabidopsis* and *Calypso* of soybean define a lineage of endogenous plant retroviruses. *Genome Res* **12**: 122-131.
- Yano, S.T., B. Panbehi, A. Das, and H.M. Laten. (2005). *Diaspora*, a large family of Ty3-gypsy retrotransposons in *Glycine max*, is an envelope-less member of an endogenous plant retrovirus lineage. *BMC Evolutionary Biology* **5:30**: 1-14.
- Zoldos, V., D. Papes, M. Cerbah, O. Panaud, M. Besendorfer, and S. Siljak-Yakovlev. (1999). Molecular-cytogenetics studies of ribosomal genes and heterochromatin reveal conserved genome organization among 11 *Quercus* species. *Theor Appl Genet* **99**: 969-977.